# Advances in Science, Technology & Engineering Systems Journal

**Mr. Abdullah El-Bayoumi**
Cairo University, Egypt

**Mr. Manu Mitra**
University of Bridgeport,
USA

**Mr. Manikant Roy**
IIT Delhi, India

**Mr. Aamir Nawaz**
Gomal University, Pakistan

# Editorial

In this issue, the journal brings together a collection of research that exemplifies the ongoing pursuit of intelligent systems that serve real human needs. The studies featured are rooted in applied innovation where technology meets societal challenges head-on. From advancing prompt engineering in artificial intelligence to enhancing assistive navigation, enabling real-time creative education, addressing linguistic inequality in language processing, and optimizing healthcare for older adults, each article contributes to a broader vision of impactful, ethically grounded technology. Together, they illustrate how research can not only push boundaries but also reshape how people live, learn, communicate, and thrive.

One article investigates the role of generative artificial intelligence and the precision required in prompt engineering to guide its outputs effectively. The discussion opens with a clear technical lineage from GANs and VAEs to the more recent Transformer models underscoring their disruptive applications in fields ranging from language to pharmaceuticals. By dissecting the mechanics and implications of prompt design, the study exposes how vague or overly constrained prompts can hinder both ethical performance and creative potential. At the same time, real-world use cases in education, finance, and medicine highlight how thoughtful prompt strategies can elevate decision-making and innovation. Emerging practices such as domain-specific fine-tuning and real-time adaptive designs are explored as promising directions, affirming the foundational role of precise, responsible input design in shaping AI's future [1].

Another contribution presents a smart solution aimed at addressing outdoor mobility challenges for the blind. The system introduced in this study goes beyond conventional navigation by factoring in psychological stress levels across multiple route options. It calculates a unique stress index using route difficulty as a foundation, ultimately selecting the least taxing path for the user. This route selection is complemented by real-time navigation support and obstacle detection, creating a feedback loop that enhances safety and ease of movement. The approach reflects a significant shift toward assistive systems that are not only technically sound but also empathetic to users' lived experiences [2].

Further research introduces an AI-powered tool designed to support photography education through automated, real-time feedback. The evaluation platform utilizes convolutional neural networks to assess images for both technical elements such as exposure and lighting and creative aspects. Trained on a dataset that combines professional standards with student work, the system delivers results with high accuracy and speed, enabling timely intervention in large or resource-limited classrooms. Initial user trials suggest that learners benefit from both skill development and increased motivation, positioning the tool as a valuable supplement to traditional instruction. As the system evolves, its framework holds potential for wider application across other visual art domains [3].

A fourth study confronts a systemic gap in natural language processing: the inability of current models to perform effectively across under-resourced languages. While NLP technologies have become sophisticated for high-resource contexts, many linguistic communities remain marginalized due to limited data, complex syntactic structures, and computational constraints. The research outlines these challenges and calls for more inclusive design strategies that support linguistic diversity and global accessibility. It emphasizes that equitable NLP systems must be grounded in culturally and structurally aware modeling, supported by policies that prioritize digital inclusion and local language preservation [4].

Another study focuses on the role of smart medical technologies in managing chronic diseases among older adults, particularly those living independently. The analysis considers how systems such as remote monitoring, AI analytics, and IoT-based devices can align with personal health needs and behavioural patterns. Using psychological and behavioural frameworks, the research outlines a methodology for designing products that accommodate both physiological requirements and adoption barriers like the digital divide. Strategies for fostering engagement and ease of use are also proposed, demonstrating that thoughtful design can significantly enhance quality of life for an ageing population while easing pressure on healthcare infrastructure [5].

These articles compiled in this issue reflect the essential shift from theoretical potential to applied, meaningful outcomes. Whether refining the mechanisms that drive generative AI, enabling more humane assistive technologies, or extending the reach of digital tools to overlooked populations, this body of work reinforces the responsibility of research to serve human contexts. These studies not only showcase technical excellence but also foreground empathy, accessibility, and equity as central pillars of technological progress.

## References:

[1]    M. Leon, "Generative Artificial Intelligence and Prompt Engineering: A Comprehensive Guide to Models, Methods, and Best Practices," Advances in Science, Technology and Engineering Systems Journal, 10(02), 01–11, 2025, doi:10.25046/aj100201.

[2]    Y. Keryakos, Y. Bou Issa, M. Salomon, A. Makhoul, "Introducing a Stress Management and Navigation System for Blind Individuals," Advances in Science, Technology and Engineering Systems Journal, 10(02), 12–27, 2025, doi:10.25046/aj100202.

[3]    S. Vorapatratorn, N. Thongsibsong, "AI-Based Photography Assessment System using Convolutional Neural Networor," Advances in Science, Technology and Engineering Systems Journal, 10(02), 28–34, 2025, doi:10.25046/aj100203.

[4]    S. Obare, K. Ogada, "A review of natural language processing techniques for under-resourced languages," Advances in Science, Technology and Engineering Systems Journal, 10(02), 35–41, 2025, doi:10.25046/aj100204.

[5]    Y. Gao, F. Layas, X. Dong, Y. Li, J. Li, "A Study of the Digital Health Management Needs of the Elderly," Advances in Science, Technology and Engineering Systems Journal, 10(02), 42–48, 2025, doi:10.25046/aj100205.

**Editor-in-chief**

**Prof. Hamid Mattiello**

# ADVANCES IN SCIENCE, TECHNOLOGY AND ENGINEERING SYSTEMS JOURNAL

## CONTENTS

ASTES

# Generative Artificial Intelligence and Prompt Engineering: A Comprehensive Guide to Models, Methods, and Best Practices

Maikel Leon[*]

*Department of Business Technology, Miami Herbert Business School, University of Miami, Miami, Florida, USA*

A R T I C L E   I N F O

A B S T R A C T

*This article enhances discussions on Generative Artificial Intelligence (GenAI) and prompt engineering by exploring critical pitfalls and industry-specific advantages. It begins with a foundational overview of AI evolution, emphasizing how generative models such as GANs, VAEs, and Transformers have revolutionized language processing, image generation, and drug discovery. Prompt engineering is highlighted as a key methodology for directing model outputs with precision and ethical awareness, enabling applications in Natural Language Processing (NLP), content personalization, and decision support. The revised sections detail how prompt engineering can be misapplied, underscoring common errors like overly restrictive or ambiguous prompts that compromise GenAI's accuracy, ethicality, and creative capacity. Equally, the paper showcases high-impact use cases in finance, education, healthcare, and beyond, illustrating how carefully formulated prompts can strengthen risk detection, enhance student learning, improve clinical decision-making, and foster product innovation. The expanded discussion of industry alignment illustrates the tangible value these techniques offer across diverse sectors, ultimately reinforcing the notion that prompt engineering is central to maximizing GenAI's transformative potential. Future directions address emerging trends, from multimodal fusion and domain-specific fine-tuning to adaptive prompt designs that leverage real-time user feedback, further solidifying the role of responsible prompt engineering in shaping the next generation of intelligent and ethically aligned AI solutions.*

## 1. Introduction

Artificial intelligence (AI) has progressed remarkably since its early stages, transitioning from systems governed by rigid rules to more adaptive, data-driven approaches capable of discerning intricate patterns [1]. Recently, the emergence of GenAI and robust prompt engineering techniques has redefined how organizations, researchers, and practitioners approach an expansive range of tasks—including, but not limited to, text analytics, visual design, product innovation, and strategic decision-making [2]. This document extensively evaluates these methodologies, presenting how GenAI fuels the creation of entirely new data. At the same time, prompt engineering directs and molds model outputs for improved precision, originality, and ethical oversight [3].

AI originated in the mid-20th century with influential figures such as Alan Turing, whose seminal inquiries into computational processes laid the groundwork for machine-mediated reasoning. During the 1950s and 1960s, there was pronounced enthusiasm for symbolic manipulation and expert systems [4], propelling significant investment and heightened aspirations for AI breakthroughs.

Over time, these expectations were periodically unmet, leading to intervals known as AI winters, marked by waning research backing and tempered academic excitement. Nonetheless, incremental advancements in ML methodologies, algorithmic efficiency, and representational frameworks persisted. By the late 1980s and 1990s, rekindled interest in neural-network-based models, coupled with improving hardware performance, catalyzed the resurgence of AI. The advent of big data analytics and deep learning architectures in the early 2010s firmly established AI as a predominant technological force, furnishing the foundations for innovative developments in GenAI and modern ML paradigms [5].

GenAI can be seen as a type of advanced Machine Learning (ML) algorithm category that produces novel yet meaningful data points by internalizing learned distributions within extensive datasets. Such approaches extend beyond conventional classification and prediction paradigms, paving the way for generating text, images, music, and complex molecular structures [6]. Prompt engineering, by contrast, constitutes a systematic process of formulating concise, context-specific prompts to orient generative models to-

---
[*]Corresponding Author: Maikel Leon (mleon@miami.edu)

ward desired behaviors. As Large Language Models (LLMs) evolve in capacity and scope, carefully devised prompts serve as a vital interface, shaping clarity, domain-specific alignment, and overall efficacy [7].

In addition to providing historical context, this paper aims to:

- Highlight gaps in existing research where more systematic experimentation on GenAI and prompt engineering is needed.
- Compare popular GenAI models at a conceptual level and discuss how they suit different tasks.
- Provide practical guidance on best practices for prompt design, with examples illustrating successful and unsuccessful outcomes.
- Expand on ethical and societal implications, focusing on the bias, potential misuse, and long-term effects of generative AI in diverse industries.

## 2. Large Language Models

LLMs represent a transformative advancement in AI, designed to comprehend, generate, and engage in human-like language interactions. With remarkable proficiency, these models leverage vast amounts of data and sophisticated algorithms to perform many language-related tasks. The development of LLMs is driven by creating more intuitive and versatile AI systems that can seamlessly integrate into various facets of business and society. By enabling machines to understand and generate natural language, LLMs bridge the gap between human communication and machine processing, facilitating more effective and meaningful interactions [8].

The capabilities of LLMs extend beyond simple text generation; they encompass understanding context, maintaining coherence over extended discourse, and adapting to diverse linguistic styles and domains. This versatility positions LLMs as essential tools for enhancing productivity, automating complex tasks, and fostering innovation across multiple industries [9]. As organizations increasingly seek to harness the power of AI to gain competitive advantages, LLMs offer a robust foundation for developing intelligent applications that can drive strategic decision-making and operational excellence.

### 2.1. History

The evolution of LLMs can be traced back to the early developments in NLP and ML. Initial attempts focused on rule-based systems and statistical methods, which laid the groundwork for more advanced models. The introduction of neural networks marked a significant milestone, enabling machines to learn from data more flexibly and on a larger scale. However, transformer architectures in the mid-2010s revolutionized the field, providing the necessary framework for building models with unprecedented capacity and performance.

The release of models such as BERT and GPT series demonstrated the potential of LLMs to perform a wide range of tasks with minimal task-specific training. These models capitalized on large-scale pre-training on diverse datasets, followed by fine-tuning for specific applications, achieving state-of-the-art results in various benchmarks [10]. The continual scaling of model parameters and training data has further enhanced the capabilities of LLMs,

enabling them to generate more coherent and contextually relevant outputs. This historical trajectory underscores the rapid advancements in computational power, data availability, and algorithmic innovations that have propelled LLMs to the forefront of AI research and application [11].

### 2.2. Architecture

The architecture of LLMs is primarily based on transformer networks, which utilize self-attention mechanisms to process and generate language. Unlike traditional recurrent neural networks, transformers can handle long-range dependencies and parallelize computations more efficiently, making them well-suited for training on extensive datasets. The core components of an LLM architecture include multiple layers of attention and feed-forward neural networks, which collectively enable the model to capture complex linguistic patterns and contextual relationships.

A typical LLM consists of an encoder and a decoder, although many modern architectures, such as the GPT series, employ only the decoder component for generative tasks. The self-attention mechanism allows the model to weigh the importance of different words in a sequence, facilitating a deeper understanding of context and meaning. Positional encoding is also incorporated to retain the order of words, which is crucial for maintaining coherence in generated text. Layer normalization, residual connections, and dropout techniques are employed to enhance training stability and prevent overfitting [3].

The scalability of LLM architectures is a key factor in their success. By increasing the number of layers, attention heads, and parameters, LLMs can achieve higher levels of performance and adaptability. This scalability is complemented by advancements in distributed computing and parallel processing, which enable the training of extremely large models on vast datasets [12]. The architectural innovations in LLMs have improved their ability to generate high-quality text and expanded their applicability to a broader range of tasks, including translation, summarization, and conversational agents.

Large-scale transformer training often incorporates gradient accumulation to handle very large batch sizes without exceeding GPU memory. Some frameworks use mixed-precision training (e.g., FP16) to reduce memory usage and speed computation. When scaling to billions of parameters, advanced optimizers like Layer-wise Adaptive Rate Scaling (LARS) can further improve convergence in deep networks. While these methods do not alter the fundamental self-attention architecture, they are critical for practical, large-scale LLM implementations.

### 2.2.1. Comparison with Other GenAI Models

Although transformers dominate many modern NLP tasks, other generative architectures retain niche advantages. RNNs and LSTMs can be more efficient for simpler tasks or smaller datasets, albeit with limitations in handling long-range context. VAEs offer interpretable latent spaces, supporting tasks like anomaly detection or data compression. Meanwhile, GANs excel in image and audio synthesis, though maintaining equilibrium between generator and discriminator can be challenging. The choice of architecture of-

ten hinges on domain constraints, data availability, and the desired trade-off between interpretability and performance.

In practical deployments, modern LLMs are typically evaluated on domain-specific tasks or industry benchmarks, such as human evaluation of text coherence, code generation accuracy, or specialized QA metrics. For instance, some organizations measure how well a large transformer-based model answers financial or legal queries compared to in-house experts or test chatbot performance on thousands of real customer interactions. These benchmarks provide pragmatic insights into how architectural differences (e.g., number of attention heads) and training optimizations translate into real-world improvements in quality and user satisfaction.

## 2.3. Applications

LLMs have found applications across various industries, leveraging their ability to understand and generate natural language to drive innovation and efficiency. In healthcare, LLMs analyze medical records, diagnose, and personalize patient care through tailored communication. By processing vast amounts of unstructured data, these models can identify patterns and insights that inform clinical decision-making and improve patient outcomes. They play a critical role in risk assessment, fraud detection, and customer service automation in the financial industry. They can analyze market trends, generate financial reports, and provide real-time support through intelligent chatbots, enhancing operational efficiency and enabling more informed investment strategies. Similarly, LLMs facilitate personalized learning experiences in the education sector by adapting educational content to individual student needs, automating grading, and providing instant feedback [13].

The realm of content creation and media has also been transformed by LLMs, which generate articles, scripts, and marketing materials with minimal human intervention. This capability accelerates content production and allows for greater customization and scalability. Additionally, LLMs enhance human-machine interactions through virtual assistants and conversational agents that engage users in meaningful and contextually relevant dialogues [14].

Beyond these sectors, LLMs are instrumental in research and development, aiding in literature reviews, hypothesis generation, and data synthesis. Their ability to process and generate language at scale makes them invaluable tools for accelerating innovation and fostering collaborative efforts across disciplines. The broad applicability of LLMs underscores their potential to drive significant advancements and create new opportunities in an increasingly data-driven and interconnected world.

## 2.4. Debrief

The widespread adoption of LLMs underscores their profound impact on technological and organizational landscapes. Their ability to process and generate human-like language has enhanced existing processes and paved the way for new applications and business models. Organizations leveraging LLMs benefit from increased efficiency, reduced operational costs, and the ability to deliver more personalized and engaging experiences to their stakeholders.

LLMs also play a pivotal role in enabling data-driven decision-making by providing deeper insights into vast and complex datasets. Their capacity to analyze unstructured data sources, such as text and speech, complements traditional data analysis methods, offering a more comprehensive understanding of market trends, customer behavior, and operational performance. This integration of LLMs into analytical frameworks empowers businesses to make more informed and strategic decisions, fostering a culture of innovation and continuous improvement [15].

Moreover, LLMs' scalability and adaptability remain relevant to evolving business needs and technological advancements. As models grow in size and complexity, their capabilities expand, allowing them to tackle more sophisticated tasks and integrate seamlessly with emerging technologies such as the Internet of Things (IoT) and augmented reality (AR). This adaptability enhances the longevity of LLM investments and ensures that organizations can stay ahead in a competitive and rapidly changing environment.

However, deploying LLMs also requires carefully considering ethical and operational challenges. Issues such as data privacy, bias in generated content, and the potential for misuse require robust governance frameworks and responsible AI practices. Addressing these challenges is essential for maximizing the benefits of LLMs while minimizing potential risks, ensuring that their integration into business processes aligns with organizational values and societal expectations.

LLMs have established themselves as indispensable tools in modern business, driving innovation and competitive advantage. Their ongoing development and integration into various sectors promise to unlock new possibilities and redefine the boundaries of what is achievable with AI. As organizations continue to navigate the complexities of digital transformation, LLMs will undoubtedly play a central role in shaping the future of work, communication, and strategic decision-making [16].

## 3. Understanding GenAI

Recognizing AI's historical trajectory and the underlying mechanisms that empower generative models can help data scientists and business innovators unlock fresh avenues for problem-solving and creative invention across multiple domains.

## 3.1. Evolution of AI: From Rule-Based to Generative Models

Early AI systems hinged on predefined logic rules and static processes. Although useful in constrained scenarios, these approaches could not adapt to subtle or evolving tasks. As the volume of digitized data swelled and computational resources advanced, ML and data-centric models gradually supplanted rigid rule-based tools [17]. Researchers came to appreciate that constructing models capable of generating new data instances and distinguishing and classifying existing data expanded the range of potential AI applications [18]. Out of this realization, GenAI emerged, providing a fertile ground for synthesizing novel text, images, or even decision-support insights. This paradigm shift spurred breakthroughs—from creating coherent language passages to rendering high-resolution images—and revolutionized industrial processes by streamlining innovation, elevating

personalized consumer services, and minimizing labor-intensive manual workflows.

## 3.2. Key GenAI Models: RNNs, LSTMs, GPT, and More

Several well-established models form the backbone of GenAI:

- Restricted Boltzmann Machines (RBMs): Early probabilistic frameworks that represent data distributions by connecting observed units with latent variables, serving as precursors to deeper generative architectures.
- Variational Autoencoders (VAEs): Leverage probabilistic encoders and decoders to map data into a latent representation, facilitating structured data generation and meaningful interpolation in a continuous space.
- Generative Adversarial Networks (GANs): Employ a two-model framework (generator and discriminator) in competitive training. The generator aims to produce highly realistic outputs, while the discriminator evaluates authenticity, thereby driving continuous improvement in generation quality [19].
- Recurrent Neural Networks (RNNs): Capture sequence dependencies through hidden states that evolve, though they often grapple with gradient-related challenges for long sequences. Despite these drawbacks, they were a fundamental step in modeling linguistic structures [20].
- Long Short-Term Memory (LSTMs): Introduce gating mechanisms to mitigate vanishing gradients, enabling more reliable handling of extended text sequences or time-dependent phenomena [21].
- Transformers (for instance, GPT): Employ attention mechanisms that operate in parallel across sequences, allowing for significantly enhanced scalability and performance in language-oriented tasks, including in-context learning and context retention over considerable text spans.

Depending on project needs—such as data volumes, output format (text, image, audio), and available computational bandwidth—each model class delivers unique strengths and may be strategically adopted for best results [22].

Although Transformers often outperform alternative GenAI architectures in large-scale text tasks, VAEs or GANs may be more suitable for image synthesis or anomaly detection. At the same time, RNNs or LSTMs can be simpler to train on smaller datasets. The best choice hinges on domain constraints and resource availability.

## 3.3. Popular Use Cases for GenAI

GenAI, by synthesizing robust and contextually appropriate data, has permeated a multitude of sectors:

- NLP: Powers automated summarization, content creation, and question-answering systems, improving customer support, knowledge dissemination, and overall operational efficiency [23].
- Image Synthesis and Editing: GANs and VAEs underpin image-to-image translation, style transfers, and photorealistic rendering, reshaping digital design and online product visualization.
- Music and Audio Generation: Sequence-based and transformer-based audio frameworks facilitate the composition of musical scores and synthetic voices, transforming entertainment and interactive voice technologies.
- Drug Discovery and Material Science: Generating new molecular and structural formulations accelerates R&D cycles, reducing the time needed for validation and optimization.
- Anomaly Detection and Pattern Recognition: Generative models model standard patterns and detect deviations, thus supporting robust fraud prevention and quality assurance initiatives.
- Data Augmentation: Generative techniques enhance predictive accuracy for various supervised learning endeavors by producing new training instances that enrich underrepresented classes [24].
- Simulation and Scenario Planning: Synthetic data fuel the simulation of market fluctuations, supply-chain constraints, or policy trade-offs, helping businesses refine their strategic planning processes [25].

Informal surveys in some organizations reveal that managers find generative models boost efficiency in drafting or analytics tasks by up to 90% in pilot projects, yet about 60% express concerns about explainability or compliance. This underscores the balance between pragmatic gains and the need for rigorous oversight in domains like insurance or healthcare [26].

## 4. Prompt Engineering

The importance of prompt engineering has risen in parallel with the widespread integration of transformer-based architectures like GPT, which respond directly to prompt instructions when generating outputs. A nuanced approach to crafting these prompts can dramatically influence model performance [27].

### 4.1. Why it Matters

Prompt engineering entails developing precise instructions—referred to as prompts—that guide generative models, specifically LLMs, toward the intended results. A well-crafted prompt can evoke succinct, context-appropriate text, structured data, or specialized solutions while minimizing irrelevant or illogical responses. As AI models expand in complexity, the prompt design stands at the forefront of practical deployment, shaping outcomes for marketing communications, medical informatics, or content moderation [28]. Effectively engineered prompts can ensure an organization's brand remains consistent, guarantee technical accuracy in specialized domains, and uphold cultural sensitivities while aligning with operational goals.

### 4.2. Prompt Types

Varying degrees of specificity characterize prompt design:

- Explicit Prompts: Clearly define content type and structure. Example: "Summarize the subsequent article in four bullet points emphasizing its principal assertions."

- Implicit Prompts: Pose more open-ended queries, leaving the model to interpret context. Example: "Reflect on the foundational ideas presented in the text."
- Creative Prompts: Intentionally inspire novel perspectives or imaginative responses. Example: "Compose a futuristic narrative inspired by the primary discoveries in this research paper."

The appropriate style depends on the degree of creative freedom desired and the degree of detail required to meet organizational or scholarly benchmarks.

## 4.3. Best Practices

The effectiveness of prompt engineering is supported by key guidelines, such as:

- Be clear and concise: Articulate instructions unambiguously to avert confusion or extraneous responses.
- Provide context: Incorporate relevant domain insights, background data, or salient references in the prompt.
- Specify the desired format: Indicate structural expectations, such as enumerated lists or succinct paragraphs.
- Encourage multiple attempts: Solicit multiple outputs or iterative feedback to refine clarity and precision.
- Balance guidance and freedom: Overly restrictive prompts may hamper creativity, while excessively broad prompts risk losing focus.
- Evaluate and iterate: Continuously refine your prompt strategies in response to model performance metrics and expert feedback.

Below are brief illustrations of how small differences in prompt design can lead to significantly different results:

Successful Prompt: "Draft a 200-word press release introducing our new data analytics platform. Emphasize speed, security, and user-friendliness. Include a short quote from the CEO."
Analysis: This prompt's clarity on style, length, and key features (speed, security, user-friendliness) helps align the generated text with the organization's marketing goals.

Unsuccessful Prompt: "Write something about our new product."
Analysis: Overly vague instructions may produce meandering or irrelevant text, failing to highlight critical selling points or match the intended brand tone.

Another Failure Example: "Provide advice for diagnosing all diseases in humans using only three bullet points."
Analysis: This is both overly ambitious and excessively constrained. It encourages the model to produce incomplete or erroneous medical advice, which poses ethical and practical risks.

## 5. Practical Applications of Prompt Engineering

Custom prompts steer AI models toward producing more accurate results but also empower these models to address complex language-based and data-driven questions spanning countless fields.

## 5.1. Improving NLP Tasks with Custom Prompts

Prompt engineering has shown substantial benefits for language-centric processes:

- Text Summarization: Prompts can delineate target length, audience, or detail level, thereby generating succinct yet comprehensive overviews.
- Sentiment Analysis: Focus model attention on emotional cues within consumer feedback, supporting targeted marketing and brand strategy.
- Text Generation: Maintain thematic continuity and organizational voice across marketing, corporate communications, or public announcements.
- Question-Answering: Embed contextual hints and clarifications in prompts to bolster factual veracity and interpretive depth.
- Text Classification: Restrict the model to specific labels or categories, improving classification consistency in legal or customer-service contexts.
- Machine Translation: Strengthen stylistic adherence and domain-specific diction in translations by offering pertinent glossaries or examples.

## 5.2. Creativity and Diversity in AI-Generated Content

Content that thrives on innovation often harnesses prompt engineering to boost ideation and novelty:

- Idea Generation: Prompt the AI to merge unrelated concepts or shift narrative points of view, expanding creative frontiers in writing or media.
- Constraint-Based Challenges: Mandate the use of specific structures, lexical elements, or rhetorical forms, fostering more unconventional outputs.
- Iterative Refinement Loops: Feed a model's output as a new prompt, encouraging sophisticated evolution of narratives, character details, or design concepts.

## 5.3. AI Ethics and Bias

Thoughtful, prompt design can serve as a bulwark against harmful outputs and biases:

- Encouraging Fairness and Inclusivity: Instruct the model to include various perspectives, effectively broadening discourse.
- Avoiding Harmful Stereotypes: Explicitly discourage hateful or derogatory content concerning respectful outcomes.
- Promoting Fact-Checked Content: Require the model to cite verifiable sources, thus curtailing misinformation and preserving credibility.

## 5.4. Personalization

Prompt engineering is pivotal for delivering tailored experiences:

- Incorporating User Preferences: Integrate a user's reading or purchase history directly into prompts for heightened personalization.

- Adjusting Language and Tone: Align the AI's outputs with brand guidelines or adapt voice and register for professional, informal, or technical contexts.
- Adaptive Learning and Tutoring: Dynamically reshape prompts in educational platforms based on each learner's prior responses, fostering individualized instruction [29].

By combining concise prompts with domain-specific data, organizations can harness AI to craft resonant, personalized communications at scale.

# 6. Improper Approaches

Misaligned or poorly structured prompts can undermine the effectiveness of even the most sophisticated GenAI models. Practitioners risk producing misleading, irrelevant, or harmful outputs by overlooking crucial clarity, context, or ethical considerations. This section investigates some of the most common pitfalls in prompt engineering, providing concrete examples and analytical commentary to illuminate why these approaches fail to achieve reliable, high-quality results [30].

## 6.1. Common Pitfalls

Poor prompt design often results from a lack of domain understanding, ambiguous phrasing, or inadequate consideration of user or organizational needs. Additionally, improper prompts can propagate undesirable biases, inaccuracies, and unproductive responses. The subsections below highlight frequent missteps, illustrating how ill-structured prompts may compromise GenAI systems' ultimate performance and trustworthiness.

## 6.2. Ambiguous Directives

One of the most frequent errors involves delivering instructions that are too broad, vague, or contradictory for the model to parse effectively. Such prompts frequently produce meandering or nonsensical outputs, undermining the project's objectives.

- Example: "Tell me something interesting."
- Example: "Explain the world in one sentence."

## 6.3. Excessive Constraints

Another improper approach involves prompts that impose stringent parameters on the model, minimizing the system's creative or inferential latitude. While clear guidance is necessary, an overly constrained prompt can stifle potentially insightful outputs.

- Example: "Answer only with exactly five words about a complex topic."
- Example: "Provide a single solution to the problem without referencing any data."

Some prompts inadvertently demand conflicting outputs or request content that cannot accurately be produced, mainly when the demands surpass model capabilities or reference non-existent data.

- Example: "Describe exactly how to cure all diseases, using five references from the future."
- Example: "Generate a precise political forecast for the next 50 years with no uncertainty."

## 6.4. Neglect of Context and Ethical Boundaries

Prompts that omit critical social, cultural, or ethical contexts can inadvertently lead to insensitive or biased outputs.

- Example: "Rank various cultures from best to worst based on your data."
- Example: "Generate a statement that supports discrimination against a particular group."

## 6.5. Debrief and Corrective Insights

The above examples illustrate how ill-conceived prompts compromise the efficacy and reliability of GenAI models. Overly broad instructions yield unfocused or convoluted content, and excessively restrictive prompts stifle the system's ability to produce nuanced, value-added information. Prompts that present contradictory or unfeasible demands provoke confusion and false claims, while a lack of ethical or cultural awareness risks perpetuating harmful stereotypes or biases [31].

To mitigate these pitfalls, practitioners should adopt a structured and mindful approach to prompt design, balancing clear directives with sufficient latitude for creativity and interpretive reasoning. Whenever possible, prompt engineers should also engage in iterative testing, monitoring, and refinement to identify and correct problematic outputs before they propagate widely. Ultimately, fostering robust prompt engineering practices and ethical oversight is pivotal for harnessing the full potential of GenAI without compromising accuracy, inclusivity, or social responsibility [26].

# 7. Challenges and Limitations of Prompt Engineering

Recognizing and addressing hurdles in prompt engineering is indispensable for creating more reliable, transparent, and equitable AI systems.

## 7.1. Limitations and Biases

Despite major strides, generative models are still prone to:

- Training Data Biases: Historical datasets may omit or underrepresent specific demographics, amplifying systemic inequalities.
- Contextual Gaps: Extended or complex queries sometimes outstrip the model's capacity to maintain accurate references.
- Unpredictable Outputs: Even meticulously designed prompts may yield unexpected results, necessitating continued vigilance in reviewing outputs.

A large-scale study revealed that about 34% of AI-generated job postings contained gendered wording that subtly favored male candidates [26], underscoring the necessity of auditing datasets

and prompts for unintended stereotypes. Additionally, adversarial "prompt injection" attempts have extracted sensitive data from generative systems in around 5% of test cases [30], highlighting the importance of monitoring and securing user-facing AI tools.

## 7.2. Balance between Guidance and Flexibility

Achieving optimal outcomes hinges on prompt specificity without hampering the model's innate innovation ability. Restrictive parameters can confine creative latitude, but an overly broad prompt can cause the system to diverge from strategic objectives. Iteration is frequently essential, with teams methodically refining prompt techniques and obtaining counsel from domain experts or stakeholders to pinpoint the intersection of control and originality.

## 7.3. Quality and Reliability in AI-Generated Content

High-caliber outputs span more than bare factual accuracy:

- Rigorous Testing and Evaluation: Employ formal metrics, human review boards, or user studies to assess the clarity, relevance, and truthfulness of outputs.
- Continuous Model Improvement: Gather real-world feedback, adapt models to new data, and fine-tune prompts to maintain high-performance standards.
- Monitoring and Maintenance: Track shifts in the model's outputs over time to detect unwanted biases or content drift.

# 8. Sectors and Industries Poised to Benefit from GenAI and Prompt Engineering

GenAI and prompt engineering offer a powerful toolkit for organizations seeking to enhance data analysis, drive product innovation, and strengthen communication strategies. These systems can substantially improve operational efficiency and unlock fresh growth opportunities by synthesizing new information and adjusting model outputs to align with specific objectives. In what follows, we examine several high-impact sectors poised to reap transformative advantages from these emerging technologies [32].

## 8.1. Financial Services and Banking

Financial institutions increasingly rely on advanced algorithms to decode intricate market signals and inform strategic decision-making. Integrating GenAI with thoughtfully constructed prompts elevates these capabilities by delivering precise, contextually relevant results.

- Automated Report Generation: Personalized investment summaries and forecasts, accelerating client services and enhancing transparency.
- Risk Assessment and Fraud Detection: Prompt-driven anomaly detection systems sift through extensive financial datasets, swiftly uncovering irregular patterns and thwarting fraudulent activities.
- Scenario Simulation: Synthetic data generation under diverse what-if prompts aids banks in evaluating market responses and regulatory changes.

- Customer Engagement: Chatbots and virtual assistants, guided by structured prompts, provide precise financial guidance at scale.

A practical prompt example in the finance sector might state: "Given two years of daily trading data for a major stock index, generate a set of five plausible volatility scenarios and highlight the potential risk factors associated with each scenario."

## 8.2. Education and Training

In educational contexts, GenAI has the potential to reshape pedagogy, course development, and student engagement. Coupled with refined prompt engineering, these solutions ensure that outputs are tailored to diverse learning levels and subject areas [33].

Below are the principal domains where GenAI and prompt engineering can advance teaching and learning:

- Customized Learning Materials: Context-sensitive prompts yield course outlines, practice problems, or revision guides, accommodating varied learner needs and abilities.
- Student Assessment: Prompts that reflect targeted cognitive skills help educators diagnose proficiency gaps and design more personalized intervention strategies.
- Intelligent Tutoring: Responsive learning assistants calibrated through precise prompts, model problem-solving processes, and offer step-by-step guidance.
- Course Content Updates: By processing recent scholarly findings, GenAI enables rapid revisions to lesson plans, ensuring instructional content is current [34].

A practical prompt example in the education sector could be: "Generate three progressively more challenging exercises on introductory algebra designed for high school students, and provide step-by-step solutions to each problem."

In a three-month pilot, a public high school leveraged an LLM-based tutor that auto-graded assignments and answered student questions. Teachers reported a 25% reduction in grading time and a 15% improvement in student engagement [26].

## 8.3. Healthcare and Medical Research

Healthcare organizations benefit substantially from GenAI and prompt engineering's capacity to improve diagnostic accuracy, accelerate drug development, and optimize various administrative functions [35].

Below are prime applications where GenAI can enrich healthcare delivery and research outcomes:

- Personalized Treatment Options: Patient-specific prompts empower generative models to simulate outcomes for alternative therapies, supporting medical practitioners in selecting the best treatment plans.
- Drug and Therapy Discovery: Automated mechanisms for molecular analysis help identify promising compounds at lower cost and in shorter timeframes than conventional methods.
- Real-Time Diagnostic Support: Prompt-driven systems evaluate patient data quickly, furnishing prioritized recommendations or alerts in high-pressure clinical settings.

- Patient Communication: Language models, directed by carefully structured prompts, generate coherent responses to frequently asked questions, reducing administrative overhead and clarifying medical guidance.

A practical prompt example in healthcare might be: "Using anonymized patient records, propose three possible treatment plans for a middle-aged patient with Type 2 diabetes and coexisting hypertension, and highlight relevant clinical guidelines."

From an overarching perspective, GenAI increases care precision and expedites key medical research tasks. By examining patient data holistically and issuing prompt-driven insights, these systems allow healthcare professionals to devote more time to complex clinical judgments and patient-centered decision-making.

## 8.4. Other High-Impact Domains

Beyond finance, education, and healthcare, various industries can harness GenAI's predictive, analytical, and creative functionalities to achieve superior efficiency and innovation [36].

Below are examples of sectors where GenAI and prompt engineering can add remarkable value:

- Manufacturing and Supply Chain: Predictive maintenance schedules and automated design prototypes can be generated via prompt-driven models, streamlining production and curbing operational bottlenecks.
- Legal and Compliance: Drafting briefs, preparing compliance documents, and reviewing regulations become more efficient when orchestrated through prompts that yield structured, domain-consistent outputs.
- E-Commerce and Retail: Highly personalized product recommendations and dynamic design ideas build more substantial customer experiences, providing real-time adaptability in a competitive marketplace.
- Marketing and Content Creation: Targeted campaigns and novel messaging formats can be generated through prompts that integrate brand narratives with pertinent consumer data.

A practical prompt example relevant to manufacturing might be: "Given a set of sensor readings from a production line over the past month, generate three potential equipment failure scenarios and propose a maintenance schedule to mitigate these risks."

The scope of GenAI's relevance spans an impressive spectrum of operations. From predictive analytics in supply chain management to data-driven marketing and retail innovation, these technologies aid decision-makers in tackling complexity, reducing uncertainties, and driving sustained organizational growth [37].

## 8.5. Unleashing Value Across Multiple Domains

In sectors ranging from financial services to healthcare and beyond, GenAI demonstrates its capacity to automate and enrich core operations, leveraging prompt engineering as a bridge between user-driven requirements and machine intelligence [38]. By carefully structuring these prompts, organizations capture the precise insights or solutions they seek, enabling wide-ranging efficiencies, faster innovation, and personalized offerings. As data-centric methodologies continue to converge, GenAI's ability to provide adaptive outputs at

scale is a pivotal catalyst for economic and societal progress. Ultimately, thorough implementation of prompt engineering practices positions businesses and institutions to thrive in today's dynamic, information-driven landscape [39].

## 9. Emerging Trends in Prompt Engineering

These forward-looking areas will likely expand GenAI's potential, growing its acceptance and strategic importance across numerous enterprises and academic domains.

### 9.1. Utilizing Advanced AI Models and Techniques

Ongoing research continues to amplify generative modeling:

- Fine-Tuning and Transfer Learning: Adapt large foundational models to specific tasks, such as legal analysis or molecular design, by integrating domain-specific layers [40].
- Multimodal AI Models: Fuse text, images, audio, and video to enrich comprehension and facilitate context-aware generation.
- Contextual AI and Memory Mechanisms: Introduce extended memory or retrieval systems that broaden a model's ability to manage complex, multi-step interactions and references.

By pursuing these developments, AI practitioners aim to engineer platforms that seamlessly navigate real-time complexities and provide actionable insights in increasingly sophisticated scenarios.

### 9.2. The Fusion of Human and AI Creativity

Collaboration between humans and advanced AI stands to shape entirely new creative spaces:

- Co-Creation and Brainstorming: Teams use AI-driven outputs as starting points or augmentations, applying professional acumen to polish and refine.
- Augmented Expertise: Specialists in medicine, law, and research can benefit from draft analyses or visual aids generated by AI, thereby improving both efficiency and quality.
- Ethical and Responsible Creativity: Ensuring content respects cultural norms and includes multiple perspectives remains critical as AI influences more creative fields.

In this paradigm, AI is a computational tool and a partner in imagining and shaping novel concepts.

### 9.3. Prompt Engineering in the AI-Driven Economy

Prompt engineering is quickly becoming a cornerstone of digital transformation and value creation:

- Business Communications: Automated drafting accelerates the production of proposals, memos, and reports, boosting organizational throughput.
- Accelerated Research and Development: Synthesizing data expedites prototyping and hypothesis testing, unlocking new opportunities across science and engineering.
- Personalized Customer Experiences: Companies cultivate stronger relationships and heightened customer loyalty by aligning outputs to individual preferences.

- Fostering Creativity and Innovation: Interactive, prompt-driven tools inspire fresh product concepts and services, bolstering long-term growth and strategic adaptability.

Astute, prompt engineering becomes a linchpin for maintaining a competitive edge as more enterprises integrate AI into their core processes.

## 10. Master Prompt Engineering with these critical tips and best practices

By adhering to practical guidelines and thoughtfully revising prompts, individuals and organizations can harness the extensive benefits of AI-generated content, mitigating common risks [41].

### 10.1. Embarking on a Journey into Prompt Engineering

- Understand your AI model's underpinnings, including architecture, training data, and known performance characteristics.
- Define concrete objectives and performance benchmarks, enabling systematic assessment of content quality and success metrics.
- Initiate with simple prompts, gradually incorporating nuanced details and complex constraints to refine outputs.
- Experiment with input style and parameter variations, collecting multiple model outputs for comparative assessment.
- Employ quantitative measures—like relevance or perplexity—and human review to evaluate clarity, factual accuracy, and creative fidelity.
- Share insights with the broader AI community, fostering a collaborative environment that accelerates best practices and innovation.
- Continuously prioritize ethics and accountability by detecting and addressing biases or misinformation that may surface.

### 10.2. Developing a Productive Workflow

1. Planning and Research: Delve into pertinent literature, existing architectures, and domain-specific challenges.
2. Objective Setting: Pinpoint the precise functions your generated content serves, from brief synopses to imaginative expansions.
3. Drafting Initial Prompts: Assemble multiple prototype prompts that reflect different degrees of specificity.
4. Testing and Evaluation: Assess the model's outputs systematically, noting clarity, functional relevance, and creativity.
5. Iteration and Refinement: Fine-tune prompts based on performance data and critique from stakeholders or subject experts.
6. Monitoring and Maintenance: Remain vigilant regarding shifts in model outputs, updating your prompts and data sources as required.
7. Collaboration and Continuous Learning: Facilitate team communication and stay current on emerging findings and tools in prompt engineering.

### 10.3. Addressing Typical Challenges

Although prompt engineering is powerful, it comes with notable obstacles:

- AI Model Constraints: Context window limits and domain mismatches may lead to oversights or inaccuracies.
- Managing Bias: Employ systematic checks, post-processing steps, or filtering to flag or mitigate biased outputs.
- Avoiding Overly Restrictive Prompts: Over-constraining the prompt can stifle the model's inventive dimension, while vague prompts can produce irrelevant or meandering text.
- Guaranteeing Content Quality: Clearly specify formatting criteria, ensure factual correctness, and regularly test model responses for consistency.
- Fostering Creativity: Utilize open-ended prompts or iterative feedback methods to overcome generative stagnation.
- Addressing Ethical Considerations: Incorporate disclaimers, implement reliability checks, and uphold societal standards and regulatory requirements.

### 10.4. Measuring the Success of Your Results

Several metrics help evaluate performance and impact:

- Content Quality (clarity, syntactical correctness, lexical richness)
- Content Novelty (degree of originality, minimal overlap with existing data)
- Content Accuracy (verified alignment with reliable sources, domain fidelity)
- User Satisfaction (positive engagement scores, survey feedback)
- Efficiency (reduced time and computational overhead)
- Adaptability (applicability to diverse tasks or evolving scenarios)

Organizations can enhance AI-driven initiatives by tracking these parameters and iterating on prompt designs to meet or surpass evolving objectives.

## 11. Conclusion

This document has examined the synergistic domains of GenAI and prompt engineering, revealing how they collectively anchor a new wave of intelligent data-processing and content-generating systems. GenAI empowers the automated production of novel outputs, spanning textual discourse, graphical elements, and molecular constructs, whereas prompt engineering channels these capabilities toward specific objectives with ethical and contextual finesse. Appreciating the historical evolution, prominent model structures, and most prevalent applications in these fields enable practitioners and scholars to optimize their AI deployments across various industrial and research contexts. As advancements continue toward broader multimodal applications, deeper contextual understanding, and more prosperous human-AI collaboration, prompt engineering remains indispensable for shaping productive, trustworthy, and forward-thinking AI solutions that address the emerging priorities of a data-driven global economy.

## 12. Future Works

Future efforts might refine formal frameworks for prompt creation, systematically linking prompt attributes to desired outcome characteristics. Another exciting direction involves the development of adaptive prompts that draw on iterative user interactions and feedback, fine-tuning a model's output dynamically in real-time. Additionally, exploring next-generation multimodal or continuous data streams could enhance generative models' versatility, enabling them to flourish in use cases involving real-time sensor data or live video feeds. Lastly, ensuring transparency, equity, and regulatory compliance in GenAI remains a priority for public confidence and ethical standards. By delving further into these research domains, the community can guide future AI systems toward excellent reliability, interpretability, and responsible innovation, facilitating breakthroughs in how societies and industries harness the power of intelligent automation.

While this paper focuses on general best practices and conceptual frameworks, future work could include direct experimental comparisons of distinct prompting strategies or GenAI architectures. Evaluating prompt styles, output quality, or error rates across multiple domains would provide more substantial evidence of each method's effectiveness and improve real-world applicability.

Furthermore, a systematic literature review could shed more light on multimodal prompt design or domain-specific fine-tuning. A deeper analysis of long-term societal implications—covering potential misuse, emergent behaviors, and ethical oversight—would strengthen preparedness for rapid AI adoption. Such expansions could broaden the reference pool to reflect contemporary sources and top-tier AI conference findings, complementing the foundational material presented here.

## Acknowledgment

## References

[1] M. León, R. Bello, K. Vanhoof, "Cognitive Maps in Transport Behavior," in 2009 Eighth Mexican International Conference on Artificial Intelligence, 179–184, IEEE, 2009, doi:10.1109/MICAI.2009.31.

[2] M. Leon, L. Mkrtchyan, B. Depaire, D. Ruan, K. Vanhoof, "Learning and clustering of fuzzy cognitive maps for travel behaviour analysis," Knowledge and Information Systems, 39(2), 435–462, 2013, doi:10.1007/s10115-013-0616-z.

[3] M. León, "Fuzzy Cognitive Maps as a Bridge Between Symbolic and Sub-Symbolic Artificial Intelligence," International Journal on Cybernetics & Informatics, 13(4), 57–75, 2024, doi:10.5121/ijci.2024.13405.

[4] M. Leon, "Aggregating Procedure for Fuzzy Cognitive Maps," The International FLAIRS Conference Proceedings, 36(1), 2023, doi:10.32473/flairs.36.133082.

[5] A. Ghimire, J. Prather, J. Edwards, "Generative AI in Education: A Study of Educators' Awareness, Sentiments, and Influencing Factors," 2024, doi:10.48550/ARXIV.2403.15586.

[6] M. León, N. M. Sánchez, Z. Z. García, R. B. Pérez, "Concept Maps Combined with Case-Based Reasoning in Order to Elaborate Intelligent Teaching/Learning Systems," in Seventh International Conference on Intelligent Systems Design and Applications (ISDA 2007), 205–210, IEEE, 2007, doi:10.1109/ISDA.2007.11.

[7] M. León, G. Nápoles, R. Bello, L. Mkrtchyan, B. Depaire, K. Vanhoof, "Tackling Travel Behaviour: An Approach Based on Fuzzy Cognitive Maps," International Journal of Computational Intelligence Systems, 6(6), 1012–1039, 2013, doi:10.1080/18756891.2013.816025.

[8] M. León, "Comparing LLMs Using a Unified Performance Ranking System," 2024, doi:10.5121/ijcsit.2023.15103.

[9] M. León, H. DeSimone, "Advancements in Explainable Artificial Intelligence for Enhanced Transparency and Interpretability across Business Applications," Advances in Science, Technology and Engineering Systems Journal, 9(5), 9–20, 2024, doi:10.25046/aj090502.

[10] M. León, "Toward the Application of the Problem-Based Learning Paradigm into the Instruction of Business Technology and Innovation," International Journal of Learning and Teaching, 10(5), 571–575, 2024, doi:10.18178/ijlt.10.5.571-575.

[11] H. DeSimone, M. León, "Leveraging Explainable AI in Business and Further," in 3rd IEEE Opportunity Research Scholars Symposium, 2024, doi:10.1109/ORSS.2024.1234567.

[12] M. León, "Harnessing Fuzzy Cognitive Maps for Advancing AI with Hybrid Interpretability and Learning Solutions," Advanced Computing: An International Journal, 15(5), 1–23, 2024, doi:10.5121/acij.2024.150501.

[13] M. León, "Generative AI as a New Paradigm for Personalized Tutoring in Modern Education," International Journal on Integrating Technology in Education, 13(3), 49–63, 2024, doi:10.5121/ijite.2024.13304.

[14] M. León, "Benchmarking Large Language Models with a Unified Performance Ranking Metric," International Journal on Foundations of Computer Science & Technology, 14(4), 15–27, 2024, doi:10.5121/ijfcst.2024.14402.

[15] M. León, "The Needed Bridge Connecting Symbolic and Sub-Symbolic AI," International Journal of Computer Science, Engineering and Information Technology, 14(1), 1–19, 2024, doi:10.5121/ijcseit.2024.14101.

[16] M. León, "Leveraging Generative AI for On-Demand Tutoring as a New Paradigm in Education," International Journal on Cybernetics & Informatics, 13(5), 17–29, 2024, doi:10.5121/ijci.2024.13502.

[17] M. León, G. Nápoles, C. Rodríguez, M. M. García, R. Bello, K. Vanhoof, "A Fuzzy Cognitive Maps Modeling, Learning and Simulation Framework for Studying Complex System," in New Challenges on Bioinspired Applications: 4th International Work-conference on the Interplay Between Natural and Artificial Computation (IWINAC 2011), 243–256, Springer Berlin Heidelberg, 2011, doi:10.1007/978-3-642-21326-7_27.

[18] G. Nopoles, M. L. Espinosa, I. Grau, K. Vanhoof, R. Bello, Fuzzy cognitive maps based models for pattern classification: Advances and challenges, volume 360, 83–98, Springer Verlag, 2018.

[19] M. León, L. Mkrtchyan, B. Depaire, D. Ruan, R. Bello, K. Vanhoof, "Learning Method Inspired on Swarm Intelligence for Fuzzy Cognitive Maps: Travel Behaviour Modelling," in Artificial Neural Networks and Machine Learning–ICANN 2012: 22nd International Conference on Artificial Neural Networks, Lausanne, Switzerland, September 11-14, 2012, Proceedings, Part I, 718–725, Springer Berlin Heidelberg, 2012, doi:10.1007/978-3-642-33269-2_90.

[20] G. Nápoles, Y. Salgueiro, I. Grau, M. Leon, "Recurrence-Aware Long-Term Cognitive Network for Explainable Pattern Classification," IEEE Transactions on Cybernetics, 53(10), 6083–6094, 2023, doi:10.1109/TCYB.2022.3142284.

[21] G. Nápoles, M. Leon, I. Grau, K. Vanhoof, "FCM Expert: Software Tool for Scenario Analysis and Pattern Classification Based on Fuzzy Cognitive Maps," International Journal on Artificial Intelligence Tools, 27(07), 1860010, 2018, doi:10.1142/S0218213018600102.

[22] M. León, B. Depaire, K. Vanhoof, "Fuzzy Cognitive Maps with Rough Concepts," in Artificial Intelligence Applications and Innovations: 9th IFIP WG 12.5 International Conference, AIAI 2013, Paphos, Cyprus, September 30– October 2, 2013, Proceedings, 527–536, Springer Berlin Heidelberg, 2013, doi:10.1007/978-3-642-41142-7_53.

[23] F. Hoitsma, A. Knoben, M. León, G. Nápoles, "Symbolic Explanation Module for Fuzzy Cognitive Map-Based Reasoning Models," in Artificial Intelligence XXXVII: 40th SGAI International Conference on Artificial Intelligence, AI 2020, Cambridge, UK, December 15–17, 2020, Proceedings, 21–34, Springer International Publishing, 2020, doi:10.1007/978-3-030-63799-6_3.

[24] M. León, G. Nápoles, M. M. García, R. Bello, K. Vanhoof, "A Revision and Experience Using Cognitive Mapping and Knowledge Engineering in Travel Behavior Sciences," Polibits, (42), 43–50, 2010, doi:10.17562/PB-42-6.

[25] H. DeSimone, M. Leon, "Explainable AI: The Quest for Transparency in Business and Beyond," in 2024 7th International Conference on Information and Computer Technologies (ICICT), IEEE, 2024, doi:10.1109/icict62343.2024.00093.

[26] M. Alier, F.-J. García-Peñalvo, J. D. Camba, "Generative Artificial Intelligence in Education: From Deceptive to Disruptive," International Journal of Interactive Multimedia and Artificial Intelligence, 8(5), 5, 2024, doi:10.9781/ijimai.2024.02.011.

[27] J. Su, W. Yang, "Unlocking the Power of ChatGPT: A Framework for Applying Generative AI in Education," ECNU Review of Education, 6(3), 355–366, 2023, doi:10.1177/20965311231168423.

[28] M. Leon, "Business Technology and Innovation Through Problem-Based Learning," in Canada International Conference on Education (CICE-2023) and World Congress on Education (WCE-2023), CICE-2023, Infonomics Society, 2023, doi:10.20533/cice.2023.0034.

[29] H. Wang, A. Tlili, R. Huang, Z. Cai, M. Li, Z. Cheng, D. Yang, M. Li, X. Zhu, C. Fei, "Examining the applications of intelligent tutoring systems in real educational contexts: A systematic literature review from the social experiment perspective," Education and Information Technologies, 28(7), 9113–9148, 2023, doi:10.1007/s10639-022-11555-x.

[30] E. A. Alasadi, C. R. Baiz, "Generative AI in Education and Research: Opportunities, Concerns, and Solutions," Journal of Chemical Education, 100(8), 2965–2971, 2023, doi:10.1021/acs.jchemed.3c00323.

[31] D. BAİDOO-ANU, L. OWUSU ANSAH, "Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning," Journal of AI, 7(1), 52–62, 2023, doi:10.61969/jai.1337500.

[32] E. Struble, M. Leon, E. Skordilis, "Intelligent Prevention of DDoS Attacks using Reinforcement Learning and Smart Contracts," The International FLAIRS Conference Proceedings, 37(1), 2024, doi:10.32473/flairs.37.1.135349.

[33] X. Zhai, X. Chu, C. S. Chai, M. S. Y. Jong, A. Istenic, M. Spector, J.-B. Liu, J. Yuan, Y. Li, "A Review of Artificial Intelligence (AI) in Education from 2010 to 2020," Complexity, 2021, 1–18, 2021, doi:10.1155/2021/8812542.

[34] C.-C. Lin, A. Y. Q. Huang, O. H. T. Lu, "Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review," Smart Learning Environments, 10(1), 2023, doi:10.1186/s40561-023-00260-y.

[35] W. Holmes, K. Porayska-Pomsta, K. Holstein, E. Sutherland, T. Baker, S. B. Shum, O. C. Santos, M. T. Rodrigo, M. Cukurova, I. I. Bittencourt, K. R. Koedinger, "Ethics of AI in Education: Towards a Community-Wide Framework," International Journal of Artificial Intelligence in Education, 32(3), 504–526, 2021, doi:10.1007/s40593-021-00239-1.

[36] K. Zhang, A. B. Aslan, "AI technologies for education: Recent research & future directions," Computers and Education: Artificial Intelligence, 2, 100025, 2021, doi:10.1016/j.caeai.2021.100025.

[37] L. Chen, P. Chen, Z. Lin, "Artificial Intelligence in Education: A Review," IEEE Access, 8, 75264–75278, 2020, doi:10.1109/ACCESS.2020.2988510.

[38] G. Nápoles, I. Grau, R. Bello, M. León, K. Vanhoof, E. Papageorgiou, "A Computational Tool for Simulation and Learning of Fuzzy Cognitive Maps," in 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 1–8, IEEE, 2015, doi:10.1109/FUZZ-IEEE.2015.7338005.

[39] G. Nápoles, J. L. Salmeron, W. Froelich, R. Falcon, M. Leon, F. Vanhoenshoven, R. Bello, K. Vanhoof, Fuzzy Cognitive Modeling: Theoretical and Practical Considerations, 77–87, Springer Singapore, 2019, doi:10.1007/978-981-13-8311-3_7.

[40] M. León, "The Escalating AI's Energy Demands and the Imperative Need for Sustainable Solutions," WSEAS Transactions on Systems, 23, 444–457, 2024, doi:10.37394/23202.2024.23.46.

[41] G. Nápoles, F. Hoitsma, A. Knoben, A. Jastrzebska, M. Leon, "Prolog-based agnostic explanation module for structured pattern classification," Information Sciences, 622, 1196–1227, 2023, doi:10.1016/j.ins.2022.12.012.

# Introducing a Stress Management and Navigation System for Blind Individuals

Youssef Keryakos[*,1], Youssef Bou Issa[1], Michel Salomon[2], Abdallah Makhoul[2]

[1]*Ticket lab, Antonine university, Baabda, Lebanon*

[2]*Femto-St, Franche-comte university, Besancon, France*

| ARTICLE INFO | ABSTRACT |
|---|---|
| | *The most challenging task in daily life of blind individuals is navigating outdoors. In this context, we are introducing and describing a navigation system that will provide two important tasks for blind individuals. Initially, the system will suggest the least stressful route for the blind to navigate among the various possible paths between a starting point and a destination. Finally, the system will provide real-time navigation guidance and real-time obstacle detection. The process to identify the less stressful route begins with the identification of all possible routes then calculating the Index of Difficulty for each Route (IDR), afterwads calculating the index of stress (ISR) based on the IDR for each route, eventually the route with lowest ISR is selected as the least stressful.* |

## 1. Introduction

Walking is a common and important mode of transportation for all individuals, but it can present challenges for those who are blind. While it may be a simple task for healthy individuals with normal vision to navigate from one point to another, blind individuals may experience difficulty, stress, and challenges in doing so. In previous studies we have identified a correlation between obstacles along a path and the level of stress experienced by blind individuals navigating that path [1]. Specifically, blind individuals navigating a road with many obstacles may experience higher levels of stress than those navigating a less obstructed road. Therefore, and since blind individuals lack visibility and cannot avoid obstacles encountered in their paths which can lead to harmful injuries, and since they cannot choose the right, less obstructed and stressful road, we are proposing a stress manipulating and navigation system to assist blind individuals in navigating more safely and managing stress. At an initial phase the proposed system will identify the least stressful route for the blind individual to travel on.

To identify the least stressful route the system will generate an index of difficulty and an index of stress for each possible route between start and destination points. By index we mean a numerical value. On the one hand the road difficulty index is a value that primarily depends on parameters specific to the road. On the other hand, the road stress index is a value that primarily depends on the index difficulty of the road among blind person specific parameters.

After identifying the least stressful route to travel on, the system will be navigating the blind individual in real time. This task will be achieved by detecting obstacles in real time using image processing techniques based on deep learning, and using voice feedback in a mobile application to alert the user to obstacles encountered and direct them along the entire path. In future work, the results of the system will be evaluated by comparing stress indices for blind users in the case of round trips or navigation on the same path or route more than once. Stress information collected during navigation by sensors integrated into the smart cane will provide feedback on a blind person's navigation experience. This feedback will enable further personalization of the road stress index for a blind person.

The remainder of the paper is organized as follows. First, Section 2 presents obstacles classification and properties. Then a state-of-the-art survey on navigation and guidance systems for blind people is made in Section 3. The concept of the proposed navigation system is introduced in Section 4, while the road difficulty index is described in Section 5. Some simulation results on the identification of the least stressful route are presented in Section 6. The penultimate Section 7 discusses the implementation of the proposal into a real navigation system and more particularly its two main steps: identifying the less stressful route and real-time navigation. In section 8, we are evaluating the navigation system Finally, Section 9 concludes the paper.

*Corresponding Author: Youssef Keryakos, Beirut, Lebanon, youssef.keryakos@ua.edu.lb

## 2. State of art for navigation and guidance system for blind people

Statistics published by the World Health Organization (WHO) have revealed that one-sixth of the world population is visually impaired, and that figure is sharply increasing. Assisting blind individuals in traveling was and still is an important research area. Different approaches, technologies and navigation tools developed or used for the blind individuals will be listed in this section. The different technologies and approaches that will be listed were grouped into 5 categories: Localization technologies, Obstacles Recognition methods, Feedback methods, Devices, and Software-based systems.

### 2.1. Localization technologies

Below we are listing some Triangulation localization techniques using different technologies and hardware.

**Triangulation using GPS-based System:** The Global Positioning System (GPS), using satellites, is mainly used for outdoor navigation. The GPS receiver determines its distance from several satellites, using data from the ephemeris that employs the triangulation principle by which the time the data are sent is compared to the time it is received. This method requires the GPS receiver to be well exposed to the satellite, which prevents indoor locations. GPS systems are widely used for outdoor navigation [2]. The main disadvantage of GPS localization is that the GPS signal strongly degrades inside buildings, between tall buildings, or in dense forest areas such as parks [3].

**Triangulation using the ZigBee-based System:** ZigBee [4] is a wireless navigation that can be used in indoor/outdoor navigation systems. It offers high location accuracy, wide coverage, simple infrastructure, low cost, fast real-time navigation and low power consumption. ZigBee Triangulation is based on IEEE 802.15.4, which uses the 868 MHz band in Europe, the 915 MHz band in North America, and the 2.4 GHz band [5].

**Triangulation using a Bluetooth-based System:** Such systems are based on the Bluetooth Triangulation [6] method, as it is an easy method already deployed in mobile phones, cameras, handheld devices, and gadgets that can be connected over a shorter distance. It is low cost, has a reduced weight and reduced size, and provides power savings. A limitation of the system is that bluetooth can only communicate over a short distance and is thus not usable for long distance communication.

**Triangulation using a Wi-Fi-based System:** A Wi-Fi-based navigation system [7] enables device connection anytime, anywhere, at home, work. Moreover, it is compatible with a wide variety of different systems. Wi-Fi has advantages over other communication systems, but still has its limits: high power consumption and it takes some time to establish a connection. For example, the authors in [8] presented a system for locating nurses, doctors or other patients within a hospital. The system can also be used to locate blind individuals. In [9], Wi-Fi triangulation has been used to help blind people get around in places close to their position.

### 2.2. Obstacles Recognition Methods

In the following, we list the different methods of obstacle recognition. Obstacles can be detected and recognized using different technologies like camera image, video processing or using sensors.

**Camera-based Image Processing Navigation System:** This system uses a camera to identify and detect mobile and static obstacles. It can identify obstacles and nearby objects in real-time. While this system has a lot of advantages, it has some limitations: high processing and power consumption [10]. In [11], the authors explained how cutting-edge machine vision methodologies offer tools to assist with basic human needs such as psychological capacity, individual movement, and activities of daily living. In [12], the system uses pre-stored images of the floor to guide people. This system finds the obstacle-free route by comparing current floor images with pre-stored images and checking for introduced obstacles.

**Ultrasonic Sensor-based Navigation Systems for Blinds:** Systems like these are based on "ultrasound waves". An ultrasonic transmitter will transmit an ultrasonic wave that will be reflected by any object or obstacle. The time of arrival of the wave added to the time reflected will identify the distance between the transmitter and the reflector. The drawback of this technology is the "short communication range", and the inclusion of sight light among obstacles and sensors is considered essential [13].

**Speech-to-Text-based Navigation System:** These systems let blind individuals translate vocal or spoken language into text. These systems use a variety of computational devices. The speech to text technology is also known as automatic speech recognition.

**Text-to-Speech-based Navigation System:** These systems render texts from documents or images into speech [14].

### 2.3. Feedback Methods

Below we are listing some methods that provide feedback to the user and warning of nearby obstacles.

**Haptic Feedback:** A cane called WeWalk uses haptic feedback and ultrasonic sensors to let blind individuals scan their surroundings in [15]. In [16], a cane uses light and ultrasonic sensors to scan surroundings and uses audio and haptic feedback/vibration to warn the user of nearby obstacles. In [17], the author presented a technology to assist the visually impaired (VI) person that uses haptic feedback. This unit detects obstacles above the knee via ultrasonic sensors. In [18], the author presented another technology for VI persons that uses haptic feedback/vibration to warn the VI person of the obstacle. In [19], the author introduced another method based on the conduction feedback. The smart cane can detect the obstacles and faces of friends and family members within 10 m. A bluetooth earpiece based on bone conduction is used to convey information about the obstacles. Haptic feedback can be divided into three categories: vibrotactile, shape-changing, and kinesthetic feedback.

### 2.4. Devices

In this section we are listing some devices that can help users navigate safely. Some of these devices provide directions for the users when navigating.

**Braille Signs Tools:** These tools can simplify the route for visually impaired, but they cannot serve as routing tools [20]. These days, numerous public regions, such as emergency clinics, railroad stations, instructive structures, entryways, lifts, and so on, are outfitted with Braille signs. These tools can help the visually impaired persons identify their area, but they cannot help them find their way.

**Smart Cane Tools:** Smart canes help blind individuals to scan their surroundings and detect nearby obstacles. In [21], a smart cane detects obstacles and alerts the blind individuals by voice using a microphone that produces sounds. For indoor usage, an innovative cane navigation system was proposed [22] that uses IoT and cloud networks. Camera, microcontrollers, and accelerometers send audio messages. A cloud service is exploited in the navigation system to assist the user in navigating from one point to another. The system was seen to be very efficient in detecting hurdles and suggesting the shortest and fastest routes to the visually impaired via a cloud-based approach.

**Voice-Operated Tools:** This outdoor voice-operated navigation system is based on G.P.S., ultrasonic sensors, and voice. This outdoor navigation system provides alerts for the current position of the users and guidance for traveling. The problem with this system is that it failed in obstacle detection and warning alerts [23]. Another navigation system uses a microcontroller to detect the obstacles and a feedback system that alerts the users about obstacles through voice and vibration [24].

**RFID-Based Map-Reading Tools:** RFID is a wireless technology that can assist and help blind individuals to navigate indoor and outdoor. Short communication range for RFID is one of the primary limitations. In [25] an indoor navigation system for blind and older adults was proposed, based on the RFID technique, to assist disabled people by offering and enabling self-navigation in indoor surroundings. The goal of creating this approach was to handle and manage interior navigation challenges while taking into consideration the accuracy and dynamics of various environments. The system was composed of two modules for navigation and localization—that is, a server and a wearable module containing a microcontroller, ultrasonic sensor, RFID, Wi-Fi module, and voice control module. The results showed 99% accuracy in experiments. The time the system takes to locate the obstacle is 0.6 s. Another map-reading system based on RFID provides solutions for visually disabled persons to pass through public places using an RFID tag grid, a bluetooth interface, a RFID cane reader, and a personal digital assistant [26]. Unfortunately, this system is costly and there is a chance of collision in heavy traffic. A map-reading system is relatively expensive because of the hardware units it includes, and its limitation is that it is unreliable for areas with heavy traffic. Another navigation system based on passive RFID, proposed in [27], is equipped with a digital compass to assist the visually impaired. The RFID transponders are mounted on the floor, as tactile paving, to build RFID networks. Localization and positioning are equipped with a digital compass, and the guiding directions are given using voice commands.

## 2.5. Software-based systems

In the following we are listing some software-based systems that help blind users to navigate safely and provide directions while navigating. Some of these softwares can also detect obstacles.

Visually impaired individuals utilize a variety of tactile tools, such as raised point maps, small-scale prototypes, and magnet boards, following orientation and mobility (O&M) training. While these multimodal maps have been proposed to assist in navigation for the blind and visually impaired, they do have limitations, including the inability to update their content. To address these limitations, accessible interactive maps have been developed as a solution [28]. In one instance, the authors of [29] employed a participatory design approach to create an augmented reality map specifically for O&M classes. This prototype incorporates projection, audio output, and the use of tactile tokens, enabling visually impaired individuals to both explore and construct maps. Another innovative approach is the SmartTactMaps introduced by [30], which leverages smartphones to assist blind individuals in exploring tactile maps. Additionally, [31] presented a 3D environmental map generated using an RGB-D sensor, capable of extracting semantic information from RGB images to aid visually impaired individuals in navigating their homes. Furthermore, a 3D Printed audiovisual tactile map named LucentMaps was proposed for the visually impaired in [32]. It aimed to simplify the integration of mobile devices with physical tactile maps. Lastly, the VizMap system [33] utilizes computer vision and crowdsourcing to gather diverse information from indoor environments. It employs on-site sighted volunteers to capture various videos, which are then semantically labeled and embedded into a reconstructed 3D model. This model serves as a queryable spatial representation of the environment.

Smartphone-based navigation solutions offer users the advantages of portability and convenience. In the end of this section we outline the various solutions introduced on the smartphone platform to assist visually impaired individuals.

NavCog3, introduced by [34], represents an indoor navigation system designed to provide turn-by-turn instructions and immediate feedback in cases of incorrect orientation. Additionally, it offers information about landmarks and some nearby points of interest. The system communicates with users through audio feedback. Another innovative approach by [35] is the PERCEPT-II application, where users receive navigation instructions to their selected destination by touching specific landmarks displayed on the mobile device. These destination spots are equipped with Near Field Communication (NFC) tags. However, it's worth noting that this system has a limitation concerning the need to install and maintain a substantial number of NFC tags.

In [36], the authors introduced a smartphone application designed to be integrated with an image recognition system, creating an assisted navigation system. Depending on network availability, the system can operate in two modes: online and offline. Upon initiation, the smartphone captures an image and transmits it to a server for processing. Deep learning algorithms [37][38] on the server are then employed to identify various obstacles. Notably, the system faces significant limitations in terms of high energy consumption and the requirement for a high-speed network connection.

The TARSIUS system, as outlined in [39], aimed to enhance users' comprehension of outdoor environments. Its components include the TARSIUS mobile app, a web server, a remote assistance centre, and bluetooth LE/iBeacon tags placed at points of interest along the streets. However, the primary challenges with the

TARSIUS system involve the cost and potential signal interferences associated with placing bluetooth beacons throughout the streets.

ENVISION, as described in [40], employs a specific method to robustly and accurately detect static and dynamic obstacles in real-time video streaming captured by an average hardware capacity smartphone. Improvements can be made to the system by enhancing the obstacle recognition and classification modules to provide target users with a better understanding of their surroundings.

The Active Vision with Human-in-the-Loop for the Visually Impaired" (ActiVis) project, developed by [41], introduces a multi-modal user interface that utilizes audio and vibration cues to convey navigational information to users. The current implementation of the ActiVis system is an android app based on a Tango device and a bone-conduction headset. Real-time adaptation to feedback parameters could enhance the system's performance.

The Tactile Wayfinder, detailed in [42], comprises a tactile belt and a Personal Digital Assistant (PDA) running a wayfinder application. This application manages location and route information, transmitting path directions to the tactile display. Vibrators in the belt provide users with navigation guidance based on these directions.

## 3. Obstacles classification and characteristics

While navigating a path a blind individual can encounter many different obstacles.

After several safe and controlled re-enacting scenarios of a visually impaired person mobilizing the experiment path [1] we have identified two types of obstacles: Moving and Motionless.

Moving Obstacles: temporary obstacles that could be moved, pedestrians and animals are considered moving obstacles. Motionless Obstacles: permanent obstacles that cannot be moved, could be electric polls, walls, or degraded concrete.

Table 1: List of most obvious and most common obstacles detected when experimenting

| Moving | Motionless |
|---|---|
| Cars | Light Poles |
| Potted Plant | Trees |
| Moving Signs | Holes |
| Animals | Bushes |
| Trash Bins | Boardwalk Edges |
| Piles Of Rock | Electric Poles |
| Motorbikes | Concrete Raised beds |
| Shop tents | Sewer Caps |
| Potato Chips Shelves | Stairs |
| Ice Cream Refrigerators | Store Decoration |
| Persons | Degraded Ground Concrete |

Listing all the moving and motionless obstacles is not possible, and the list that came out of our experiment is definitely not every single obstacle found on that street as shown in Table 1. A person can throw a glass bottle on the ground and that would lead to a creation of a new obstacle, which we did not specifically encounter but it is most likely to happen. Furthermore, we needed to go deeper with the classification. based on the state of art and our obstacles

listed above we have extracted the characteristics that obstacles share. The most critical characteristics that obstacles share is: size, position, risk, flexibility, velocity and sound.

In [22] a list of accident have been registered with blind individuals while navigation a road: Accidents with motorbikes and bicycles that rides on sidewalks, falling from stairs, falling into gaps between buses and their platform, bumping into building scaffolds and glass doors, Accidents with ironware and pipes transported over vehicles roof, Accidents with tree branches, telephone booths and charity boxes, bumping into objects along footpath, Accidents with passing bicycles, Accidents with passing vehicles, Accidents with parked automobiles, falling into holes, falling into pits, falling into underground rails.

In [43] various barriers were reported by blind individuals that impacted their safety: lag of precise location for a building entrance, unknown stairs and roadside holes, as well as obstacles above waist level, sidewalks and crossings, sidewalks often occupied by booths and bikes. Also, Public transport stops lacking audio information and traffic lights without audio output have been reported as an unsafe barrier by blind individuals.

In [44], a women reported the barriers that she have encountered: motorbikes parked on the sidewalks, potted plants, holes in the sidewalk, stairs, and uneven ground and she said she can't count how often she had hurt her arm by bumping into motorbike handles, or hit her head or chin on something, or fallen into an uncovered ditch. She calls some of these sidewalks "death stations.". Also, a 36-year-old blind man said that his foot became stuck in a hole in the sidewalk, he fell and had surgery and still can't fully bend his leg.

In [45] they have identified that blind individuals may use sound to navigate and identify types of obstacles: auditory changes in the walking surface, the sounds of flowing water, and the touch of breeze and sunshine on the skin helped in understanding the changes in the environment and served as navigation cues.

## 4. The concept of our proposed navigation system

In real life context, if a blind individual is trying to navigate between two geographical points, he must adapt between many possible routes to achieve his destination. After choosing a route he will begin to parade towards his destination, whether guided or not. An early step of many steps in our navigation system is to propose for the blind individuals the less stressful route among the many possible routes to take between two points. The concept of -our proposed navigation system is as follows:

In a first step the blind individual will use a mobile application to choose vocally two geographical points at start and destination positions, the system will then identify the possible routes to take between these two points and propose vocally to the blind individual the less stressful route. To identify the less stressful route the system will calculate the index of difficulty for each possible route (IDR) and then select the route that has the lesser IDR value. After proposing the less stressful route the blind individual should adopt it and his traveling journey will start.

In step 2 the blind individual will begin to walk to his destination, while walking the system will identify and detect in real time the obstacles encountered by the blind individuals. notification and

alerts will be forwarded vocally. The system will identify the type and the position of the obstacle.

# 5. Road difficulty index (IDR)

IDR depends on:

- sum of index of difficulty for each obstacle (IDO)

- total number of obstacles on the road

- road total length ($x_l$)

- road slope ($x_s$)

- accessibility of the road ($x_a$)

- average number of obstacles per meter of road ($x_n$)

$$\text{IDR} = \frac{\sum \text{IDO}}{1000 \cdot x_n \cdot x_l} + x_l + \frac{|x_s|}{10} + x_w + x_a + \frac{x_n}{2} \quad (1)$$

$$1 \le IDR \le 10 \quad (2)$$

## 5.1. IDR variables with their potential values

Each obstacle depends on six variables: size, position, risk, flexibility, velocity and sound. Each variable has been associated to a ponderation coefficient as shown in Table 2:

Table 2: Obstacle Variables

| Size | Position | Risk | Flexibility | Velocity | Sound |
|------|----------|------|-------------|----------|-------|
| s | p | r | f | v | m |
| 2/24 | 5/24 | 8/24 | 3/24 | 4/24 | 2/24 |

$$\text{IDO} = \frac{1}{24}(2s + 5p + 8r + 3f + 4v + 2m) \quad (3)$$

$$1 \le IDO \le 5 \quad (4)$$

Obstacles can have different sizes, human beings perceive size of objects compared to his size thus size of obstacle is identified as small, normal or large. Values for variables are respectively +1 for a small obstacle, +3 for a normal obstacle and +5 for a large obstacle.

An obstacle can have a position compared to ground. Position can be assessed into two main categories above or below ground. For above ground position can be also assessed into two sub categories below waist or above waist. An obstacle above the ground below waist can be detected much easier than an obstacle below ground or above waist. An obstacle below ground can be detected with a cane much easier than an obstacle above waist. Therefore, values of variable p are respectively +1 for an obstacle above ground below waist, +3 for an obstacle below ground and +5 for an obstacle above ground above waist.

Obstacles when encountered can cause risks for humans. An obstacle can be not risky at all, an obstacle can cause small injuries and finally an obstacle can cause serious injuries. Therefore, values of variable r are respectively +1 for not causing risks, +3 for small

injuries and +5 for serious injuries. An obstacle can present three levels of flexibility. An obstacle can be flexible, stiff and medium. A stiff obstacle can harm the blind and cause injuries, blind will be stressed in front of this kind of obstacle and will try to estimate the size of the obstacle to avoid it. Values of variable f are respectively +1 for a flexible obstacle, +3 for a medium obstacle and +5 for a stiff obstacle.

An obstacle can be stationary (fix) or have the capability to move. Moving obstacles can have low speeds or high speeds. A fixed object compared to a moving obstacle can be easily detected by blind person's cane. Furthermore, low speeds compared to high speed obstacles can be easily detected by a blind person, although low speed obstacles can harm the blind but are less dangerous than high speed obstacles. Therefore, values of variable v are respectively +1 for fixed obstacles, +3 for moving obstacles with low speeds and +5 for moving obstacles with high speeds.

An obstacle can be a source of sound emission. The sound emitted by obstacles can be subbed into two categories: familiar sounds or unfamiliar ones. Obstacles with no sound are the hardest to detect because a blind person can use sounds to redirect and estimate the type of obstacles near him. Unfamiliar sounds will make the blind stressed but the blind can estimate or try to locate the source of the sound. Obstacles with familiar sounds are the easiest to detect or locate nearby (like car sounds, bike sounds, sound of human talking...). Therefore, values of variable m are respectively +1 for an obstacle emitting familiar sounds, +3 for an obstacle emitting unfamiliar sounds and +5 for an obstacle not emitting sounds.

**Example:**
**IDO calculation for an obstacle with the below characteristics:**
1. Large obstacle in size: $s \rightarrow +5$
2. Position is below ground: $p \rightarrow +3$
3. Small injuries in risks : $r \rightarrow +3$
4. Flexibility is stiff: $f \rightarrow +5$
5. Stationary obstacle: $v \rightarrow +1$
6. No sounds emission: $m \rightarrow +5$

$$\text{IDO} = \frac{1}{24}(2 \cdot 5 + 5 \cdot 3 + 8 \cdot 3 + 3 \cdot 5 + 4 \cdot 1 + 2 \cdot 5) \quad (5)$$

$$\text{IDO} = \frac{78}{24} = 3.25 \quad (6)$$

Roads are travelled by a blind person from a starting point to a destination. The distance between a starting point and a destination point is usually measured in meters and designated as road length. Therefore, road total length ($x_l$) is equal to road length/1000.

A road can be horizontally straight or can present slopes.by road we mean from the starting point to the destination point of the blind user or routes that are parts of a road from start to destination where each route can be considered as a road in itself. Therefore, road slope ($x_s$) is the value of the road slope between 0 and 100%.

Accessibility for blind people may vary from road to road and country to country. Accessibility of a road for a blind person involves ensuring that individuals with visual impairments can navigate and travel safely and independently. Here are some criteria that makes a road accessible for blind users: clear sidewalks and

pathways, tactile paving, audible signals at crosswalks, tactile crosswalks,.... Therefore, accessibility of the road for blind people ($x_a$) values are respectively 0 if the road is well accessible for a blind person, 0.5 if the road presents a medium accessibility and 0.75 if the road is not at all accessible for blind.

when traveling roads weather conditions may vary. Therefore, values of weather conditions variable ($x_w$) are respectively 0 if no rain, 0.5 if rainy and 0.75 if snowy.

Finally, the last variable that IDR of a road depends on is average number of obstacles per meter of road ($x_n$). Therefore, value of $x_n$ can be between 0 and maximum 2 obstacles/m (we considered that if this value is bigger than 2 the blind user would be sucked and cannot walk or move anymore).

**Example:**
**IDR calculation for a road with the below characteristics:**

- Road total length: $\rightarrow x_l$=250 m /1000 = 0.25

- Road slope: $\rightarrow x_s$=5

- Weather conditions rainy: $\rightarrow x_w$=0.5

- Accessibility of the road not at all: $\rightarrow x_a$=0.75

- Number of obstacles: $\rightarrow x_n$=0.04

- Total number of obstacles: $\rightarrow$ 10

- IDO = 1, 1.2, 3, 3.75, 2, 2.35, 1.3, 4.2, 1.1, 1.4

- $\sum$ IDO = 21.3

$$\text{IDR} = \frac{21.3}{10} + 0.25 + 0.5 + 0.5 + 0.75 + 0.02 \quad (7)$$

$$\text{IDR} = 4.15 \quad (8)$$

### 5.1.1. ISR variables with their potential values:

ISR depends on different parameters:

- if blind person is used to the road ($i_1$)

- if the blind person is traveling alone or having assistance ($i_2$)

- if the blind person is having assistance from any hardware ($i_3$)

- Stress initial status of the blind ($i_4$,$i_5$,$i_6$,$i_7$)

$$\text{ISR} = \text{IDR} \times (i_n) \quad (9)$$

$$\text{ISR} = \text{IDR} \times i_1 \times i_2 \times i_3 \times i_4 \times i_5 \times i_6 \times i_7 \quad (10)$$

$$0.015526 \le \text{ISR} \le 10 \quad (11)$$

Values of $i_1$ are respectively +1 if the blind person is traveling the road for the first time, +0.75 if the blind has traveled the road a few times and +0.5 if the road is very well known.

Values of $i_2$ are respectively +1 if the blind is traveling the road alone, +0.5 if he is traveling assisted by a dog and +0.25 if he is assisted by a person.

Values of $i_3$ are respectively 1 if the user is not using either the cane or GPS, 0.75 if he is using any of the cane or GPS and 0.5 if he is using both of the cane and GPS.

$i_4$ represents the value of fatigue, $i_4$ values range between 0.5 and 1. $i_4$ depends on three parameters: time of experimentation ($i_t$), age of the blind person ($i_a$) and on earlier injuries ($i_i$). $i_t$ values are respectively +1 if experimentation is being held in the afternoon, 0.75 if it is being held in the noon and 0.5 if in the morning. $i_a$ values are respectively +1 if user age is greater or equal than 60, 0.5 if user age is greater or equal than 30 and 0.25 if user age is lesser than 30.

$i_i$ values are respectively +1 if user presents earlier injuries in both hands and legs, 0.5 if user presents injuries in hands only or legs only and 0.25 if the user doesn't have injuries at all.

$$i_4 = \frac{i_t + i_a + i_i}{3} \quad (12)$$

$i_5$ represents a value for the initial heart beat rate (bpm) at the starting point of the road, $i_5$ values range between 0.5 and 1. Normal heart rate values range between 60 and 100 so we will consider that between this range $i_5$ value will be equal to 0.5. the maximum bpm for any person is related to his age.

$$\text{max\_bpm} = 220 - \text{age} \quad (13)$$

Therefore $i_5$ value will be equal to 1 if initial heart beat rate was bigger or equal than max bpm. If initially a blind person has a heart rate of x:

$$x > 100 \quad (14)$$

and

$$x < \text{max\_bpm} \quad (15)$$

$i_5$ value will be equal to:

$$i_5 = 0.5 + \frac{0.5}{\text{max\_bpm} - 100} \cdot (x - 100) \quad (16)$$

$i_6$ represents a value for the initial EDA at the starting point of a road, $i_6$ values range between 0.5 and 1. normal EDA values collected with prototypes are distributed or range between 12 and 515 For an initial EDA value at the starting point of the road lesser or equal to 100 $i_6$ will be equal to 0.5. For an initial EDA value at the starting point of the road bigger or equal to 500 $i_6$ will be equal to 1. if an initial EDA value at the starting point of the road x is in between 100 and 500 $i_6$ will be equal to:

$$i_6 = 0.5 + \frac{(x - 100) \cdot 0.5}{400} \quad (17)$$

$i_7$ represents a value for the initial EMG at the starting point of a road, $i_7$ values range between 0.5 and 1. normal EMG values collected with prototypes are distributed or range between 0 and 1000 For an initial EMG value at the starting point of the road lesser or equal to 200 $i_7$ will be equal to 0.5. For an initial EDA value at the starting point of the road bigger or equal to 800 $i_7$ will be equal to 1 . if an initial EMG value at the starting point of the road x is in between 200 and 800 $i_7$ will be equal to:

$$i_7 = 0.5 + \frac{(x - 200) \cdot 0.5}{600} \tag{18}$$

## 6. Simulating the identification of the least stressful route

Any blind person can choose to travel from a starting point to a destination point. Same destination can be reached using different roads. Each road will have its own index of difficulty and its own index of stress. Our system will propose for the blind which road or segments to navigate based on the indexes. In this section we will describe the concept of our proposed navigation system by listing the details of two simulations. First simulation was taking the example of a blind user trying to navigate from university of Toulouse to cathedral of St Etienne. The second simulation was done on our experimental path in Lebanon. There is a notable difference between the two simulations. In the first simulation no ISRs were calculated since we don't have any recorded bio signal values for any blind user (ISR is related to parameters specific for a blind user such as heart beat, EDA, EMG,,...)

### 6.1. Simulating the identification of the least stressful route on the experimental path

The annotated image is taken from google maps is shown in figure 1. It represents multiple roads that a blind person can take from the University of Toulouse, France to Cathedral Saint-Etienne. We are considering that the blind person is traveling from starting point to the destination point in a very good weather condition (no rain). Furthermore, all routes correspond to the same geometrical area with 155-meter elevation above sea level but presents no slopes and all the roads present none accessibility for blind persons. Roads can have different lengths and different widths and different types.

All possible intersections are represented by Alphabetical letters with A the starting point (university of Toulouse). To start with the blind person has two possible road options AB or AC.

Based on the IDR values our navigation system will propose for the blind person the road which have the minimum IDR(AB) as hown in Table 3.When the blind is at point B, he has two possible roads that he can take to achieve his destination: BD or BE. Based on the IDR values shown in Table 4, our navigation system will propose for the blind person the road BD which has the minimum IDR.
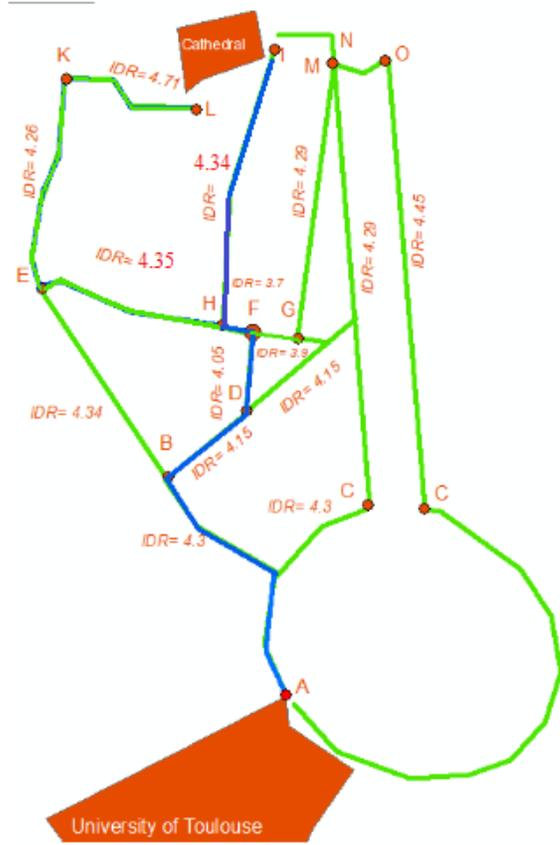


Figure 1: Image showing available routes with their IDRs from university of toulouse to Cathédrale St Etienne.

Table 3: Calculating the IDR for routes AB and AC

| Variable | AC | AB |
|---|---|---|
| $x_l$ | 0.247 | 0.24 |
| $x_a$ | 0.75 | 0.75 |
| $x_s$ | 0 | 0 |
| $x_w$ | 0 | 0 |
| Total number of obstacles detected | 68 | 70 |
| $x_n$ | 0.27 | 0.29 |
| $\sum$ IDO | 217 | 222 |
| IDR | 4.32 | 4.3 |

Table 4: Calculating the IDR for routes BD and BE

| Variable | BD | BE |
|---|---|---|
| $x_l$ | 0.073 | 0.205 |
| $x_a$ | 0.75 | 0.75 |
| $x_s$ | 0 | 0 |
| $x_w$ | 0 | 0 |
| Total number of obstacles detected | 28 | 36 |
| $x_n$ | 0.38 | 0.17 |
| $\sum$ IDO | 88 | 119 |
| IDR | 4.15 | 4.34 |

From point D, the blind person will be redirected into DF, because as can be seen on Figure 1 the route DF has the lowest IDR value. When the blind is at point F, he has two possible roads that he can take to achieve his destination FH or FG.

Table 5 shows that FH has the lowest IDR value and thus it is this road that will be advised to the blind person. Then, starting from point H, the less stressful road according to our approach is HI since, as shown in Table 6, it has less detected obstacles.

Table 5: Calculating the IDR for routes FH and FG

| Variable | FH | FG |
|---|---|---|
| $x_l$ | 0.014 | 0.04 |
| $x_a$ | 0.75 | 0.75 |
| $x_s$ | 0 | 0 |
| $x_w$ | 0 | 0 |
| Total number of obstacles detected | 5 | 10 |
| $x_n$ | 0.36 | 0.25 |
| $\sum$ IDO | 14 | 30 |
| IDR | 3.7 | 3.9 |

Table 6: Calculating the IDR for routes HI and HE

| Variable | HI | HE |
|---|---|---|
| $x_l$ | 0.23 | 0.18 |
| $x_a$ | 0.75 | 0.75 |
| $x_s$ | 0 | 0 |
| $x_w$ | 0 | 0 |
| Total number of obstacles detected | 54 | 49 |
| $x_n$ | 0.23 | 0.27 |
| $\sum$ IDO | 175.5 | 161 |
| IDR | 4.34 | 4.35 |

## 6.2. Simulating the identification of the least stressful route on the experimental path

Experiments were conducted on a road that has a total length of 800 meters as shown in Figure 2.

The road was partitioned into 3 equivalent sectors in length. Each of the sectors has a length of 265 meters. Each of the sectors has a well-known start and destination point and can be considered as a single road in itself. Sector 1 is represented with the line that connects A with B, sector 2 represented by the line that connects B to C, sector 3 represented by the line that connects C to D and sector-redirection represented by the line that connects C to E. Table 7 shows the IDR variables for each sector, Table 8 shows the IDR calculation for each sector.

### 6.2.1. ISR calculation for each sector for different experimentations

Each of the experimentation above listed in the table is done by a different user. So in total the Table 9 show experimentations done by 5 users.



Figure 2: Image showing the experimental path with point A starting point and D and E as destination points.

Table 7: Identifying IDR variables for each sector

| Variable | Sector 1 | Sector 2 | Sector 3 | Sector redirection |
|---|---|---|---|---|
| $x_l$ | 0.265 | 0.265 | 0.265 | 0.265 |
| $x_a$ | 0.75 | 0.75 | 0.75 | 0.75 |
| $x_s$ | 0 | 10 | 35 | 20 |
| $x_w$ | 0 | 0 | 0 | 0 |
| Total number of obstacles detected | 30 | 49 | 38 | 16 |
| $x_n$ | 0.11 | 0.18 | 0.14 | 0.06 |
| $\sum$ IDO | 101 | 164 | 128 | 50 |

Table 8: IDR for each sector

| IDR | Sector 1 | Sector 2 | Sector 3 | Sector redirection |
|---|---|---|---|---|
| | 4.1 | 5.1 | 7.6 | 6 |

Table 9: ISR for each sector

| Experimentation | Sector 1 | Sector 2 | Sector 3 | Sector redirection |
|---|---|---|---|---|
| Exp 1 | 0.17 | 0.18 | 0.28 | 0.23 |
| Exp 2 | 0.16 | 0.25 | 0.42 | 0.34 |
| Exp 3 | 0.27 | 0.36 | 0.49 | 0.39 |
| Exp 4 | 0.19 | 0.32 | 0.51 | 0.38 |
| Exp 5 | 0.29 | 0.31 | 0.55 | 0.43 |

# 7. Implementing the concept into a real navigation system

The proposed navigation system is designed to accomplish two primary tasks. In the initial phase, the system suggests the least stressful route for blind individuals. The user begins by selecting a starting point and a destination, prompting the system to identify all possible routes and recommend the least stressful one.

In a subsequent stage, the system provides real-time navigation for the visually impaired individual. It detects obstacles encountered during navigation and, additionally, offer real-time voice alerts to inform the user about the type of obstacle encountered. Moreover, the system redirects the user along the entire path using voice instructions to ensure a smooth and stress-free navigation experience. The navigation system was first developed as a web application and then developed as a mobile application to allow users with visual impairment to have the ability to engage with the system through voice commands, and all communication from the system to the user are conveyed through voice feedback.

## 7.1. Identification of the least stressful route

The process initiates when the user opts to navigate from a chosen starting point to a designated destination. Upon selecting these points, the system analyses all potential routes connecting the two locations and determine the least stressful option. The present outcome of this procedure is to recognize and suggest the least stressful route for the user to navigate. Currently, this process is implemented as a web application using .NET and C#. Although, the developed web application is user-friendly for individuals without visual impairments, it may pose challenges for those who are blind. this process was also integrated into a mobile application and operated by blind individuals through voice commands. We will list below the different steps of the process lifecycle.

### 7.1.1. Choosing a navigation path

In the first step the user should decide on a navigation path. We will consider as an example that the user has chosen to navigate between the University of Toulouse and the Cathedral of Saint-Etienne in Toulouse, France. After choosing the navigation path a user can choose either to automatically get all possible roads to take between the starting point (university of Toulouse) and the destination point (Cathedral of Saint Etienne) or to manually precise each intersection that possibly could be taken in the chosen navigation path.

To get automatically all possible roads that can be taken between two points: the user should first redirect to the automated system view, after redirecting the user should indicate by clicking on the displayed google maps in the view the starting point and destination point. After indicating the starting and destination points the user clicks on the get Routes button Figure 3 and 4.

Possible roads that can be taken will be displayed on google maps. Figure 5 shows 3 possible roads to take between the university of Toulouse and the cathedral of Saint-Etienne. Furthermore, each possible road will be indexed and saved to SQL Database along with all longitude and latitude points that reside on this road as shown in Table 10 and 11.
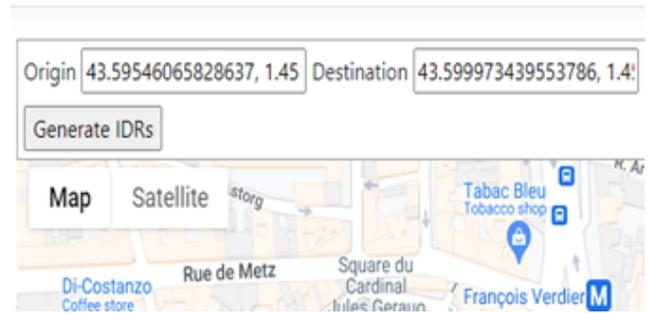


Figure 3: Image showing coordination's of origin and destination points chosen by the user.
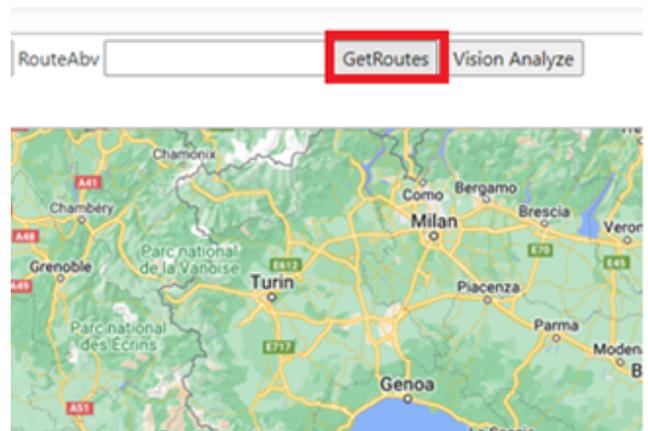


Figure 4: Image showing GetRoutes button that user clicks after choosing the start and destination points.
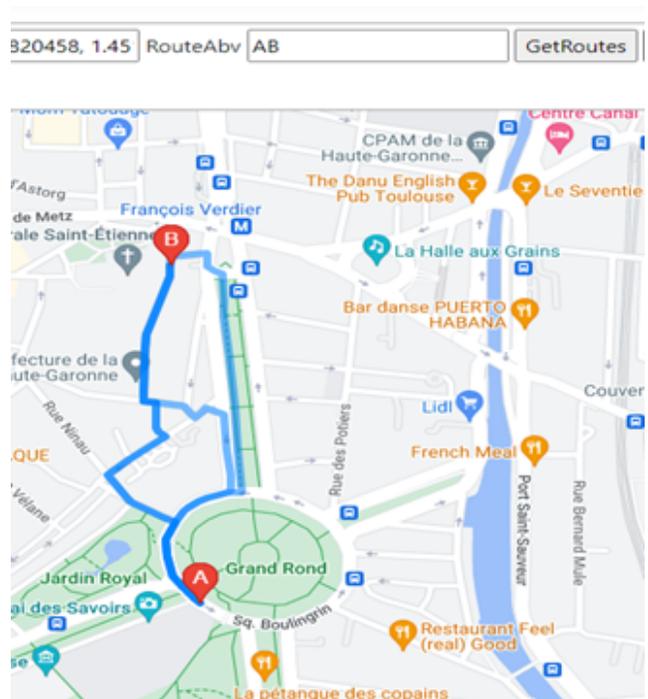


Figure 5: Image showing all possible routes displayed in google maps after clicking on getRoutes button.

Table 10: All possible routes saved into database.

| Id | Route |
|----|-------|
| 27 | Tone-0 |
| 28 | Tone-1 |
| 29 | Tone-2 |

Table 11: All geographical points for a route saved into database.

| Id | Longitude | Latitude | Route |
|-----|-----------|----------|-------|
| 507 | 1.45124 | 43.59539 | 27 |
| 508 | 1.45139 | 43.59538 | 27 |
| 509 | 1.45125 | 43.59552 | 27 |
| 510 | 1.45116 | 43.59579 | 27 |
| 511 | 1.45114 | 43.59602 | 27 |
| 512 | 1.45128 | 43.59628 | 27 |
| 513 | 1.45141 | 43.59643 | 27 |
| 514 | 1.45166 | 43.59657 | 27 |
| 515 | 1.45211 | 43.59669 | 27 |
| 516 | 1.45231 | 43.59669 | 27 |
| 517 | 1.45231 | 43.59674 | 27 |
| 518 | 1.45221 | 43.59790 | 27 |
| 519 | 1.45204 | 43.59934 | 27 |
| 520 | 1.45201 | 43.59971 | 27 |
| 521 | 1.45201 | 43.59978 | 27 |
| 522 | 1.45196 | 43.59980 | 27 |
| 523 | 1.45189 | 43.59983 | 27 |
| 524 | 1.45178 | 43.59989 | 27 |
| 525 | 1.45179 | 43.59992 | 27 |

Users can also choose to manually indicate on the displayed google maps the possible roads that exist on the chosen navigation path As shown in Figure 6. He can choose to indicate on the map all possible two points that can be possibly taken for the same intersection As shown in Figure 7.
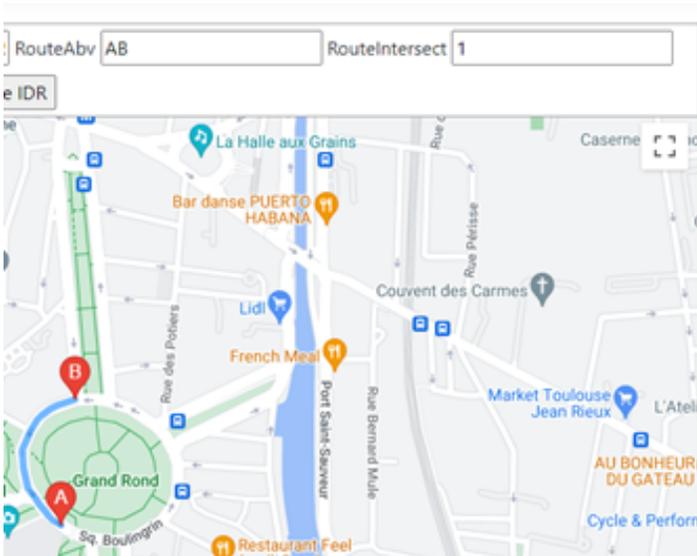


Figure 6: Image showing a manually selected route by the user.
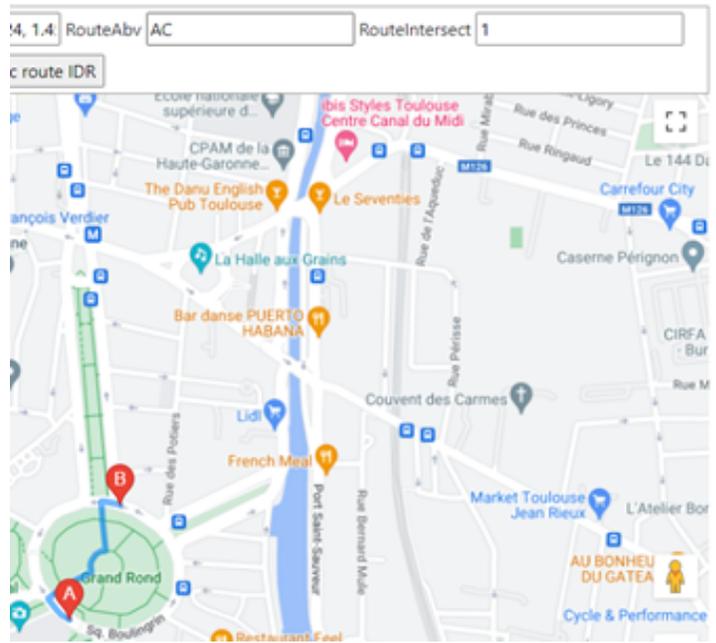


Figure 7: Image showing a manually selected possible route for the same intersection in picture above by the user.

Each possible road for each intersection will be saved to SQL database along with all the longitude and latitude points that reside on this road.

### 7.1.2. Analyzing the routes

After storing the routes and the geographical points associated with each route in the SQL database, it is time now to analyze these routes. At first, we are calculating the total distance in meters for each route. To calculate the total distance for each route we are using Google maps Api. The total distance for each route is the sum of distances between two consecutive geographical points that reside in this route. After calculating distances for each route it is time now to identify the obstacles that a blind individual can encounter on these routes. To identify the obstacles related to each route we are using 3 Apis Google Maps, Google vision and Mapillary.

The Google Maps Static Api returns an image (either GIF, PNG or JPEG) in response to an HTTP request via a URL. For each request, we can specify the location of the map, the size of the image, the zoom level, the type of map, and the placement of optional markers at locations on the map. The Google Api offers powerful pre-trained machine learning models through REST and RPC Apis. Assign labels to images and quickly classify them into millions of predefined categories. Detect objects, read printed and handwritten text, and build valuable metadata into an image catalog. The Mapillary Api allows image extraction with object detection in bounding boxes.

Using Google maps Api and Mapillary we are extracting images associated with each geographical point that resides in a route knowing its longitude and latitude values. Using Google vision and Mapillary we are identifying the labels and objects for each image associated with a point that resides in a route. Each object or label extracted will be identified stored in the database further more will be associated an IDO as shown in Table 12.

Table 12: Detected objects or labels with their IDOS in the route indexed as tone-0

| Name | Route | ApiUsed | IDO |
|------|-------|---------|-----|
| Car | Tone-0 | GoogleVision | 52 |
| Building | Tone-0 | GoogleVision | 88 |
| Car | Tone-0 | GoogleVision | 52 |
| Car | Tone-0 | GoogleVision | 52 |
| Wheel | Tone-0 | GoogleVision | 52 |
| Wheel | Tone-0 | GoogleVision | 52 |
| Car | Tone-0 | GoogleVision | 52 |
| Wheel | Tone-0 | GoogleVision | 52 |
| Car | Tone-0 | GoogleVision | 52 |
| Car | Tone-0 | GoogleVision | 52 |
| Person | Tone-0 | GoogleVision | 0 |
| Van | Tone-0 | GoogleVision | 0 |
| Bus | Tone-0 | GoogleVision | 52 |
| Bus | Tone-0 | GoogleVision | 52 |
| Bus | Tone-0 | GoogleVision | 52 |
| Bus | Tone-0 | GoogleVision | 52 |
| Building | Tone-0 | GoogleVision | 88 |
| Door | Tone-0 | GoogleVision | 0 |
| Car | Tone-0 | GoogleVision | 52 |
| Car | Tone-0 | GoogleVision | 52 |
| Building | Tone-0 | GoogleVision | 88 |
| Building | Tone-0 | GoogleVision | 88 |
| Furniture | Tone-0 | GoogleVision | 0 |
| Stop sign | Tone-0 | GoogleVision | 0 |
| Building | Tone-0 | GoogleVision | 88 |
| Person | Tone-0 | GoogleVision | 0 |
| Car | Tone-0 | GoogleVision | 52 |

### 7.1.3. Generating IDR

After storing the routes in the Database along with their geographical points and after extracting and identifying objects associated with each route using Google Maps Api, Google vision Api and Mapillary Api it is Time now to generate for each route an IDR value.

To generate an IDR value for each route the system will apply the IDR formula.

$$\text{IDR} = \frac{\sum \text{IDO}}{1000 \cdot x_n \cdot x_l} + x_l + \frac{|x_s|}{10} + x_w + x_a + \frac{x_n}{2} \quad (19)$$

$$1 \leq \text{IDR} \leq 10 \quad (20)$$

- $x_a$ is equal to 0.75 for the accessibility of the road (following our example starting from the University of Toulouse to Cathedral of Saint-Etienne).

- $x_s$ is equal to 0 for the route slope.

- $x_w$ is equal to 0, considering that the example was conducted in good weather conditions.

- $x_l$ is equal to the route total length divided by 1000.

- $x_n$ is the total count of all obstacles detected on the route divided by the route total length.

The system will be able to count the number of objects and labels associated with each route by retrieving from the database the labels and objects related to a route Id. The system will also calculate the total sum of IDOs associated with each object or labelled related to a route Id. The total length of each road is calculated by google maps API as stated before. After applying the IDR formula for each route as shown in Table 13, the system will then propose on the map the road with the least IDR for the user As shown in Figure 8.

Table 13: IDR generated and the length for each route.

| Id | Route | IDR | Length |
|----|-------|-----|--------|
| 27 | Tone-0 | 3.363 | 613 |
| 28 | Tone-1 | 3.381 | 631 |
| 29 | Tone-2 | 4.438 | 688 |
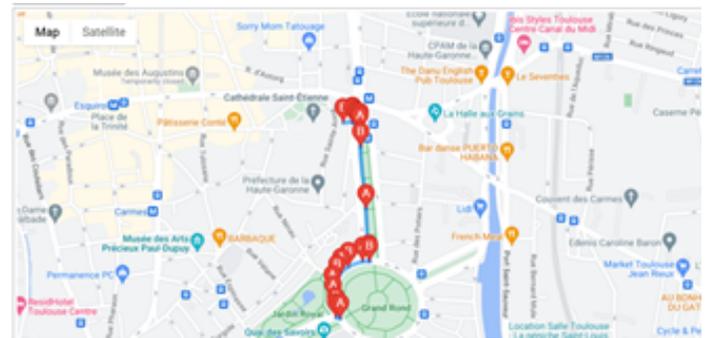


Figure 8: Image showing the generated route that has the least IDR value 3.363 .

### 7.2. Real-time navigation

This process comes after identifying the least stressful route. After proposing the least stressful route the blind user should begin to navigate into the proposed route. This process allows the user to navigate safely by getting directions via voice commands along the path and by identifying and detecting in real time the obstacles that can be encountered along the path.

### 7.2.1. Directing the blind individuals

Google maps Api provides for the blind user real-time directions while navigating the path. users will be directed by voice through the mobile application. Directions will be provided by the Google Maps Api

### 7.2.2. real-time obstacle detection

We used the built-in camera in the mobile along with image processing and machine learning techniques to identify and detect obstacles that can be encountered along the path. We are still researching this part in order to implement the best technique that will let us be able to detect and identify the majority of obstacles that can be encountered by a blind individual when navigating a route.

As for now we have been able to implement SSD [46] (Single shot detection) real time object detection with real time video capturing within a mobile application. SSD Is an object detector released

in 2017, designed for mobile and embedded vision applications. This architecture uses proven depth-wise separable convolutions to build lightweight deep neural networks. The mobile application uses TensorFlow [47] lite implementation along with our embedded custom SSD trained model.

Our custom SSD model was trained on Google Colab [48] to be able to identify between two object classes: chairs and tables. Images for these two classes were extracted from different sources from the internet in order to acquire a big image set and be able to split the set into train and test sets in order to train the model accurately.

Our custom model works well for the two classes chairs and tables As shown in Figure 9, but can identify only these two classes. To be able to identify more obstacles it should be trained on different classes which means acquiring for each new class to be detected images related to this class and furthermore acquiring a big image set which will be time consuming but can be done if no other option is available. Due to these limitations we are using a default detection model trained on the COCO dataset that can detect between 70 to 90 classes.



Figure 9: Image showing a detected chair with the custom SSD model.

## 8. Evaluation

The evaluation of our navigation system is a critical step in validating its functionality. This part is devoted to evaluating and testing the navigation system to make sure it satisfies user needs and intended goals. This section will outline the assessment procedure that was used, from collecting data sets to talking about the findings and conclusions. By doing a comprehensive analysis, we are able to see any possible problems, get useful input, and make informed decisions that will improve the prototype's usability and efficacy. Our ultimate objective is to improve and optimize the prototype in order to better meet user requirements and project objectives.

### 8.1. Overview of the navigation system used in the experimentation

The navigation system discussed earlier, has been integrated into a mobile application to enhance accessibility for blind users. The original web application was not ergonomically suited for blind individuals, making the mobile adaptation necessary. This mobile application will primarily perform the functions outlined and described earlier.

First, the application will enable blind users to interact using voice commands to select their starting location and destination. Next, it will calculate the least stressful route. The application will then provide real-time voice-guided navigation and detect obstacles along the route. Additionally, biosignals related to stress will be continuously monitored and recorded in a Google Firebase database throughout the navigation process.

### 8.2. Experimentation protocol

The experiment involved navigating a route from a residential area to the municipality's public building in the center of the town of Jezzine (the same starting point and destination as for contributions 1 and 2).

The route's infrastructure is not accessible and is not user-friendly for blind users. Six blindfolded people conducted the experiment using the navigation system. Most of the users were between 20 and 35 years old. The users were all active, healthy, physically fit, and able to walk a path on their legs, not to mention that none of the users had a heart condition. The experiments were conducted in daylight and in cool weather. For all the experiments, all the users were in a resting state before the start of the test. The preparation time before the start of the test was recorded as was the duration of the entire test for each user. The average time for a user to complete the path and the entire test was around 25 minutes.

The least stressful route among all possible routes from the starting to destination point in the experimentation was chosen by the navigation system for all the users. The least stressful route chosen, show in Figure 10, was then divided into three sectors. Bio signals data gathered for each user were then grouped by sectors. Furthermore, each user has been asked to note his stress level within a range from 0 (no stress) to 10 (the most stressed) in each sector three times: at the beginning of each sector, in the center of the sector and at the end.

For each user, data stored in the Firebase database are initially extracted in JSON format. Subsequently, this JSON data is converted to CSV format using Python. This conversion facilitates proper manipulation, visualization, and analysis of the data. Each user's CSV dataset may contain over 1000 records, with the number of records varying based on the total duration, in seconds, that the user spent walking the experimental path. The total number of records collected for all 6 users is 7892. The data samples, partitioned by sectors, are distributed as follows:

- Sector 1: 1497

- Sector 2: 3566

- Sector 3: 2829

Figure 10: Image showing the least stressful route chosen by the navigation system.

## 8.3. Discussion

The heart rate, skin conductance and muscle activity values of all users were added together and filtered by sector. Then the mean for each of the measures was calculated for each sector. The obtained values are shown in Tables 14, 15, and 16.

Table 14: Mean heart beats per minute for each sector.

|  | Sector 1 | Sector 2 | Sector 3 |
|---|---|---|---|
| **Heart Beats (Means)** | 93 | 97 | 99 |

Table 15: Mean skin conductance values for each sector.

|  | Sector 1 | Sector 2 | Sector 3 |
|---|---|---|---|
| **Skin Conductance (Means)** | 303.5 | 348.1 | 358.1 |

Table 16: Mean muscle activity values for each sector.

|  | Sector 1 | Sector 2 | Sector 3 |
|---|---|---|---|
| **Muscle Activity (Means)** | 447.3 | 485.5 | 485.47 |

A notable difference exists between the sectors concerning the heart beat, skin conductance, and muscle activity values, where we can notice that the lower means are in Sector 1, then in Sector 2 and the higher values in Sector 3. Additionally, as stated before, each user has auto-reported his stress level at the beginning, the center and the end of each sector. For each sector and for each user the highest values of stress reported was taken and added together. Next, the stress level was averaged for each sector, resulting in the values shown in Table 17. Clearly, the evolution of heart beat, skin conductance and muscle activity values are linked to the stress level. Both can accurately reflect the level of stress induced by the navigation path.

Table 17: Average stress levels for each sector

|  | Stress level (mean) |
|---|---|
| Sector 1 | 1.3 |
| Sector 2 | 3.57 |
| Sector 3 | 4.5 |

In our experiment with the full navigation system, we analyzed a route selected by the system as the least stressful, which was divided into three sectors. By incorporating self-reported stress levels from each user, we observed that all physiological measures (heart rate, skin conductivity, and muscle activity) were influenced by the distance to obstacles. This finding corroborates the results from our previous experiments described in [1]. If we consider the Tables 18 up to 21 and the corresponding Figures 11 up to 14, the following observations can be made. Specifically, users exhibited the lowest values for all biosignal measures and reported the least stress in Sector 1. Both biosignal measures and self-reported stress levels increased progressively from Sector 1 to Sector 2 and from Sector 2 to Sector 3.

Notably, despite the sectors being of equal distance, Sector 1 was traversed the fastest by all users. The variations in physiological measures, self-reported stress levels, and travel times were associated with the sector characteristics, such as obstacle density and traffic level. The path was relatively short, approximately 800 meters in length. Sectors 1 and 2 were straight, while Sector 3 was a slope that required minimal physical effort. Additionally, four out of the six users were between 20 and 30 years old.

Compared to the results from the experiment described in our second contribution [1], and considering that users traveled from the same starting point to the same destination, this experiment shows the lowest mean values for all stress-related physiological signals and the lowest reported stress levels across all three sectors. This indicates that the navigation system was effective in reducing stress for blind individuals during navigation from the starting point to the destination point.
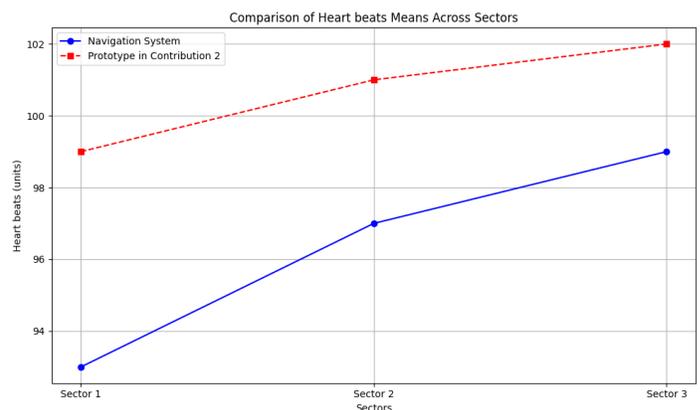


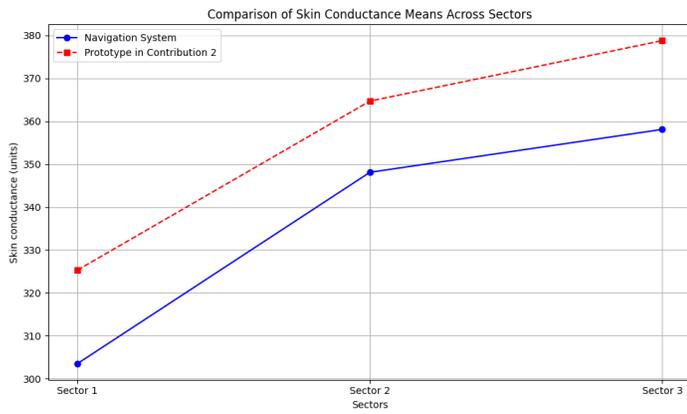Figure 11: Comparison of heart beats means across sectors.

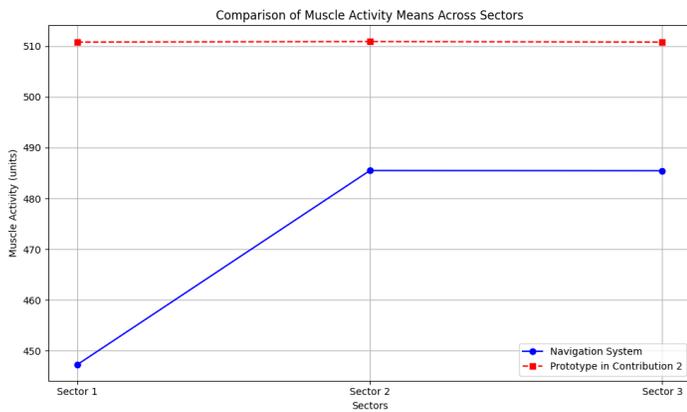Figure 12: Comparison of skin conductance means across sectors.



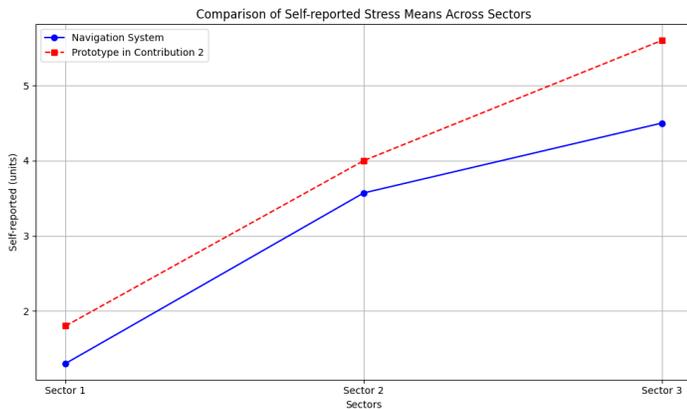Figure 13: Comparison of muscle activity means across sectors.



Figure 14: Comparison of self-reported stress means across sectors.

Table 18: Comparison of heart beats means across sectors.

|  | Sector 1 | Sector 2 | Sector 3 |
|---|---|---|---|
| **Using Navigation System** | 93 | 97 | 99 |
| **Using Prototype in Contribution 2 [1]** | 99 | 101 | 102 |

Table 19: Comparison of skin conductance means across sectors.

|  | Sector 1 | Sector 2 | Sector 3 |
|---|---|---|---|
| **Using Navigation System** | 303.5 | 348.1 | 358.1 |
| **Using Prototype in Contribution 2 [1]** | 325.3 | 364.7 | 378.8 |

Table 20: Comparison of muscle activity means across sectors.

|  | Sector 1 | Sector 2 | Sector 3 |
|---|---|---|---|
| **Using Navigation System** | 447.3 | 485.5 | 485.47 |
| **Using Prototype in Contribution 2 [1]** | 510.8 | 510.9 | 510.8 |

Table 21: Comparison of self-reported stress means across sectors.

|  | Sector 1 | Sector 2 | Sector 3 |
|---|---|---|---|
| **Using Navigation System** | 1.3 | 3.57 | 4.5 |
| **Using Prototype in Contribution 2 [1]** | 1.8 | 4 | 5.6 |

## 9. Conclusion

We have proposed a navigation system to help blind individuals to navigate safely and with less stress. Our navigation system performs two main tasks. Firstly, the navigation system suggests to the blind person the least stressful route to choose from the many possible routes between two points. Secondly, using image processing and machine learning techniques, the system detects in real time any obstacles that the blind person may encounter while navigating the route. In addition, the system provides real-time guidance to the blind person as he navigates the route.

The first process or task for the proposed system was implemented as a web application later on integrated in a mobile application. The proposed system can identify flawlessly any possible routes to take between two points and then propose for the user the route with the minimum IDR (least stressful route). Furthermore, and for future works blind users' feedback will be taken into consideration while identifying the least stressful route.

The second process, real-time object detection, was also implemented in a mobile application. We developed and trained our custom SSD model and tested it for two types of obstacle (a chair and a table). The custom model was able to accurately detect both classes. Our research continues into the second process, with the aim of building a model capable of detecting the majority of obstacles that a user may encounter.

**Conflict of Interest**    The authors declare no conflict of interest.

## References

[1] Y. Keryakos, Y. B. Issa, M. Salomon, A. Makhoul, "Correlation Between Types of Obstacles and Stress Level of Blind People in Outdoor Navigation," in Proceedings of the International Wireless Communications and Mo-

bile Computing Conference (IWCMC), Marrakech, Morocco, 2023, online: https://hal.archives-ouvertes.fr/hal-04224786.

[2] A. J. Ramadhan, "Wearable Smart System for Visually Impaired People," Sensors, **18**(3), 843, 2018, doi:10.3390/s18030843.

[3] R. K. Katzschmann, B. Araki, D. Rus, "Safe Local Navigation for Visually Impaired Users with a Time-of-Flight and Haptic Feedback Device," IEEE Transactions on Neural Systems and Rehabilitation Engineering, **26**(3), 583–593, 2018, doi:10.1109/TNSRE.2018.2800665.

[4] G. Goncalves, H. Helena, "Indoor Location System Using ZigBee Technology," in Proceedings of the 2009 Third International Conference on Sensor Technologies and Applications, 152–157, 2009, doi:10.1109/SENSORCOMM.2009.35.

[5] J. Karchňák, D. Šimšík, B. Jobbágy, D. Onofrejová, "Feasibility Evaluation of Wearable Sensors for Homecare Systems," Acta Mechanica Slovaca, **19**(2), 58–62, 2015, doi:10.21496/ams.2015.016.

[6] D. Freitas, G. Kouroupetroglou, "Speech Technologies for Blind and Low Vision Persons," Technology and Disability, **20**(2), 135–156, 2008, doi:10.3233/TAD-2008-20208.

[7] C.-H. Lim, Y. Wan, B.-P. Ng, C.-M. S. See, "A Real-Time Indoor WiFi Localization System Utilizing Smart Antennas," IEEE Transactions on Consumer Electronics, **53**(2), 618–622, 2007, doi:10.1109/TCE.2007.381746.

[8] W. Elmannai, K. Elleithy, "Sensor-Based Assistive Devices for Visually-Impaired People: Current Status, Challenges, and Future Directions," Sensors, **17**(3), 565, 2017, doi:10.3390/s17030565.

[9] R. Kanan, O. Elhassan, "A Combined Batteryless Radio and WiFi Indoor Positioning for Hospital Nursing," Journal of Communication and Software Systems, **12**(1), 34–39, 2016, doi:10.24138/jcomss.v12i1.95.

[10] J. P. Gomes, J. P. Sousa, C. R. Cunha, E. P. Morais, "An Indoor Navigation Architecture Using Variable Data Sources for Blind and Visually Impaired Persons," in Proceedings of the 13th Iberian Conference on Information Systems and Technologies (CISTI), 2018, doi:10.23919/CISTI.2018.8399264.

[11] L. Marco, G. Farinella, Computer Vision for Assistive Healthcare, Elsevier Science, Amsterdam, The Netherlands, 2018.

[12] A. Caldini, M. Fanfani, C. Colombo, "Smartphone-Based Obstacle Detection for the Visually Impaired," in Image Analysis and Processing – ICIAP 2015, volume 9279 of *Lecture Notes in Computer Science*, 480–488, Springer International Publishing, 2015, doi:10.1007/978-3-319-23231-7_43.

[13] H.-C. Wang, R. K. Katzschmann, S. Teng, B. Araki, L. Giarre, D. Rus, "Enabling Independent Navigation for Visually Impaired People through a Wearable Vision-Based Feedback System," in Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), 6533–6540, 2017, doi:10.1109/ICRA.2017.7989772.

[14] R. Luo, X. Tan, R. Wang, T. Qin, J. Li, S. Zhao, E. Chen, T.-Y. Liu, "LightSpeech: Lightweight and Fast Text to Speech with Neural Architecture Search," in Proceedings of the ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 5699–5703, 2021, doi:10.1109/ICASSP39728.2021.9414856.

[15] A. Karkar, S. Al-Maadeed, "Mobile Assistive Technologies for Visual Impaired Users: A Survey," in Proceedings of the 2018 International Conference on Computer and Applications (ICCA), 427–433, 2018, doi:10.1109/COMAPP.2018.8460195.

[16] S. Phung, M. Le, A. Bouzerdoum, "Pedestrian lane detection in unstructured scenes for assistive navigation," Computer Vision and Image Understanding, **149**, 186–196, 2016, doi:10.1016/j.cviu.2016.01.011.

[17] L. Singh, P. Mazumder, G. Sharma, "Comparison of drug susceptibility pattern of Mycobacterium tuberculosis assayed by MODS (Microscopic-observation drug-susceptibility) with that of PM (proportion method) from clinical isolates of North East India," IOSR Journal of Pharmacy, **4**, 1–6, 2014, doi:10.9790/3013-04020201-06.

[18] P. Bossé, "A Plant Identification Game," American Biology Teacher, **39**, 115, 1977, doi:10.2307/4445817.

[19] G. Sainarayanan, R. Nagarajan, S. Yaacob, "Fuzzy image processing scheme for autonomous navigation of human blind," Applied Soft Computing, **7**, 257–264, 2007, doi:10.1016/j.asoc.2005.06.005.

[20] National Instruments, "mySmartCane: Giving Freedom to Visually Impaired People," 2018, accessed August 2018.

[21] V. Mustonen, M. Pantzar, "Tracking Social Rhythms of the Heart: From Dataism to Art," Approaching Religion, **3**(2), 16–21, 2013, doi:10.30664/ar.67512.

[22] A. Riazi, F. Riazi, R. Yoosfi, F. Bahmeei, "Outdoor difficulties experienced by a group of visually impaired Iranian people," Journal of Current Ophthalmology, **28**, 2016, doi:10.1016/j.joco.2016.04.002.

[23] I. F. B. Hairuman, O.-M. Foong, "OCR Signage Recognition with Skew and Slant Correction for Visually Impaired People," in Proceedings of the 2011 11th International Conference on Hybrid Intelligent Systems (HIS), 306–310, 2011, doi:10.1109/HIS.2011.6122118.

[24] M. D. Messaoudi, B.-A. J. Menelas, H. Mcheick, "Autonomous Smart White Cane Navigation System for Indoor Usage," Technologies, **8**(3), 37, 2020, doi:10.3390/technologies8030037.

[25] J. Bai, D. Liu, G. Su, Z. Fu, "A Cloud and Vision-Based Navigation System Used for Blind People," in Proceedings of the 2017 International Conference on Artificial Intelligence, Automation and Control Technologies (AIACT 17), 416–420, 2017, doi:10.2991/aiact-17.2017.93.

[26] O. Oladayo, "A Multidimensional Walking Aid for Visually Impaired Using Ultrasonic Sensors Network with Voice Guidance," Int. J. Intell. Syst. Appl., **6**, 53–59, 2014, doi:10.5815/ijisa.2014.08.06.

[27] C. Barberis, A. Andrea, G. Giovanni, M. Paolo, "Experiencing Indoor Navigation on Mobile Devices," IT Prof., **16**, 50–57, 2013, doi:10.1109/MITP.2013.54.

[28] J. Ducasse, A. M. Brock, C. Jouffrais, "Accessible interactive maps for visually impaired users," in Mobility of Visually Impaired People, 537–584, Springer, 2018.

[29] J. Albouys-Perrois, J. Laviole, C. Briant, A. M. Brock, "Towards a Multisensory Augmented Reality Map for Blind and Low Vision People: A Participatory Design Approach," in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 629:1–629:14, 2018, doi:10.1145/3173574.3174203.

[30] T. Götzelmann, K. Winkler, "SmartTactMaps: A Smartphone-Based Approach to Support Blind Persons in Exploring Tactile Maps," in Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments, 2:1–2:8, 2015, doi:10.1145/2769493.2769497.

[31] Q. Liu, R. Li, H. Hu, D. bao Gu, "Building Semantic Maps for Blind People to Navigate at Home," in 2016 8th Computer Science and Electronic Engineering Conference (CEEC), 12–17, IEEE, 2016, doi:10.1109/CEEC.2016.7835901.

[32] T. Götzelmann, "LucentMaps: 3D Printed Audiovisual Tactile Maps for Blind and Visually Impaired People," in Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility, 81–90, 2016, doi:10.1145/2982142.2982163.

[33] C. Gleason, A. Guo, G. Laput, K. Kitani, J. P. Bigham, "VizMap: Accessible Visual Information through Crowdsourced Map Reconstruction," in Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility, 273–274, 2016, doi:10.1145/2982142.2982180.

[34] D. Sato, U. Oh, K. Naito, H. Takagi, K. Kitani, C. Asakawa, "NavCog3: An Evaluation of a Smartphone-Based Blind Indoor Navigation Assistant with Semantic Features in a Large-Scale Environment," in Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '17), 270–279, 2017, doi:10.1145/3132525.3132535.

[35] A. Ganz, J. M. Schafer, Y. Tao, C. Wilson, M. Robertson, "PERCEPT-II: Smartphone Based Indoor Navigation System for the Blind," in Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 3662–3665, 2014, doi:10.1109/EMBC.2014.6944417.

[36] B.-S. Lin, C.-C. Lee, P.-Y. Chiang, "Simple Smartphone-Based Guiding System for Visually Impaired People," Sensors, **17**(6), 1371, 2017, doi:10.3390/s17061371.

[37] S. Ren, K. He, R. Girshick, J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in Advances in Neural Information Processing Systems 28 (NIPS 2015), 91–99, 2015.

[38] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779–788, 2016, doi:10.1109/CVPR.2016.91.

[39] T. V. Mataró, F. Masulli, S. Rovetta, A. Cabri, C. Traverso, E. Capris, S. Torretta, "An Assistive Mobile System Supporting Blind and Visual Impaired People When Outdoors," in Proceedings of the IEEE 3rd International Forum on Research and Technologies for Society and Industry (RTSI), 1–6, 2017, doi:10.1109/RTSI.2017.8065886.

[40] J. C. Lock, G. Cielniak, N. Bellotto, "A Portable Navigation System with an Adaptive Multimodal Interface for the Blind," in Proceedings of the AAAI 2017 Spring Symposium on Designing the User Experience of Machine Learning Systems, 395–400, 2017.

[41] W. Heuten, N. Henze, S. Boll, M. Pielot, "Tactile Wayfinder: A Non-Visual Support System for Wayfinding," in Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges, 172–181, ACM, 2008, doi:10.1145/1463160.1463179.

[42] R. Manduchi, J. Coughlan, "(Computer) Vision Without Sight," Communications of the ACM, **55**(1), 96–104, 2012, doi:10.1145/2063176.2063200.

[43] "A Survey: Outdoor Mobility Experiences by the Visually Impaired," No further details available.

[44] "Obstacles for blind people," https://iranhumanrights.org/2018/06/drr-obstacles-for-blind-people/, 2018, accessed: 2023-08-28.

[45] M. Bandukda, A. Singh, N. Berthouze, C. Holloway, "Understanding Experiences of Blind Individuals in Outdoor Nature," in Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, CHI EA '19, 1–6, Association for Computing Machinery, New York, NY, USA, 2019, doi:10.1145/3290607.3313008.

[46] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. Berg, "SSD: Single Shot MultiBox Detector," in Proceedings of the European Conference on Computer Vision (ECCV), 21–37, 2016, doi:10.1007/978-3-319-46448-0_2.

[47] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, "TensorFlow: A System for Large-Scale Machine Learning," in Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), 265–283, 2016.

[48] E. Bisong, "Google Colaboratory," in Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners, 59–64, Springer, 2019, doi:10.1007/978-1-4842-4470-8_7.

# AI-Based Photography Assessment System using Convolutional Neural Networks

Surapol Vorapatratorn,*, Nontawat Thongsibsong

*Center of Excellence in AI and Emerging Technologies, School of Applied Digital Technology, Mae Fah Luang University, Chiang Rai, 57100, Thailand*

ARTICLE INFO

ABSTRACT

*Providing timely and meaningful feedback in photography education is challenging, particularly in large classes where manual assessment can delay skill development. This paper presents M-Stock, an AI-based automated photo evaluation system that uses Convolutional Neural Networks (CNNs) to assess student photography assignments on web browser. M-Stock evaluates both technical aspects (such as lighting, composition, and exposure) and creative elements, providing students with real-time, formative feedback. The system was trained on a diverse dataset, including student submissions and commercial standards, achieving an overall accuracy of 97.18% with an average prediction speed of 46.1 milliseconds per image. Experiments assessed the system's performance across varying resolutions and batch sizes, confirming its scalability and suitability for real-time classroom use. Additionally, a pilot study with students indicated that M-Stock's feedback positively impacted their technical skills and encouraged self-directed learning. The results demonstrate M-Stock's potential as a transformative tool for photography education, combining high accuracy, immediate feedback, and pedagogical value to support continuous learning. Future improvements will focus on refining creative assessments and expanding the system's applicability to other visual arts disciplines.*

## 1. Introduction

In recent years, digital technology has revolutionized the way photography is taught, offering students unprecedented access to resources and tools for developing their skills. University-level courses on photography increasingly emphasize both theoretical knowledge and practical expertise, aiming to produce competent professionals equipped for the rapidly evolving media and creative industries [1]. However, as photography courses expand in scope and enrolment, especially in digital classrooms, educators face significant challenges in efficiently assessing student work [2]. The task of providing timely, meaningful feedback is often hindered by the volume of student submissions, which can delay the developmental process of photography skills [3].

Traditional assessment methods for photography assignments are often manual and time-consuming, leading to delays that can impede learning and limit student engagement.

Studies have highlighted that real-time feedback plays a significant role in accelerating skill acquisition in domains requiring both technical precision and creative expression [4]. Given this, automated assessment systems powered by artificial intelligence (AI) have emerged as promising tools for enhancing the learning experience. AI technologies, especially deep learning, have shown considerable potential in automating visual assessments, enabling more personalized, consistent, and timely feedback for students [5].

Despite these advancements, current AI-based assessment systems in photography education primarily focus on evaluating technical attributes, such as lighting, composition, and exposure, often overlooking the creative and subjective aspects critical to artistic development. Furthermore, many existing tools provide only summative feedback, offering a one-time evaluation rather than iterative feedback that supports continuous learning and improvement. Addressing these gaps requires an assessment platform that can balance both technical and creative evaluations while also offering formative, actionable feedback.

*Corresponding Author: Surapol Vorapatratorn, Mae Fah Luang University,
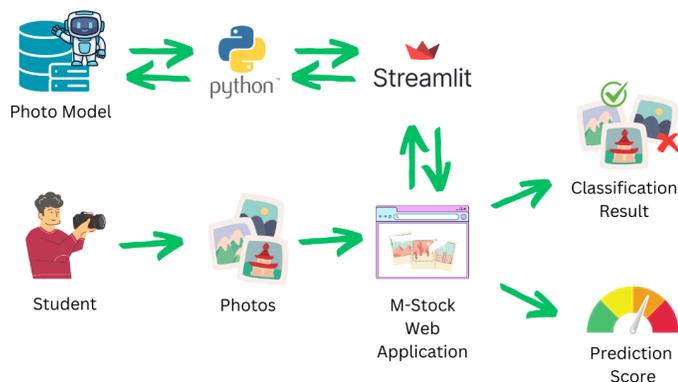Surapol.vor@mfu.ac.th

Figure 1: Overall structure of our proposed system

This paper introduces M-Stock (Mae Fah Luang University Photo Stock), an AI-driven automated photo evaluation platform designed to support student learning in photography by providing real-time feedback on both technical and artistic elements of their work. Using Convolutional Neural Networks (CNNs), M-Stock evaluates photographs based on predefined criteria developed in consultation with industry standards and educational experts, thus ensuring both relevance to the professional field and pedagogical value. In addition, M-Stock is built with scalability and ease of use in mind, allowing seamless integration into classroom environments where students can receive immediate feedback on their submissions. Overall structure of our proposed system as depicted in Figure 1.

## 2. Related Work

The integration of artificial intelligence (AI) in education has shown promising potential to enhance learning outcomes by providing personalized and adaptive feedback across various fields. In creative education, AI-driven assessment tools have been increasingly applied, yet challenges remain, particularly in domains like photography, where both technical and creative competencies are essential. This section reviews recent advancements in AI-supported educational systems, focusing on automated assessment in creative disciplines and identifying key gaps that the M-Stock system aims to address.

### 2.1. AI in Education for Automated Assessment

AI technologies, particularly deep learning, have transformed educational assessment by enabling automated grading and personalized feedback systems. These systems have proven effective in evaluating diverse student outputs, including essays, problem-solving exercises, and visual projects, providing more timely feedback than traditional methods. Adaptive learning environments and intelligent tutoring systems use AI to tailor educational content and assessment to individual learners' needs, which has been shown to improve learning efficiency and engagement [6]. Furthermore, as 21st-century learning frameworks emphasize critical thinking, creativity, and lifelong learning [7], AI-based assessments must evolve to support these skills, especially in creative subjects like photography. Furthermore, AI technologies have been applied in the context of stock photography. Platforms such as Shutterstock and Adobe Stock have incorporated AI algorithms to evaluate the quality of images submitted by photographers, offering real-time feedback and ensuring that only images meeting commercial standards are

accepted [8]. This use of AI for large-scale image evaluation highlights its potential for integration into photography education, where it can be used to assess student submissions and provide immediate feedback on technical aspects such as focus, lighting, and composition [9]. However, most existing systems in this category are optimized for structured and quantifiable tasks, such as quizzes and assignments that focus on objective metrics. This approach is limited in addressing subjective assessments, such as those required in photography education, which involve creative expression and aesthetic judgment.

### 2.2. Automated Assessment in Photography Education

In photography education, AI-based assessment tools have typically focused on evaluating technical attributes, such as exposure, sharpness, and composition. As professional photography requires both technical proficiency and artistic expression, it is critical that educational tools reflect industry standards and expectations [10]. Recent research by [11], has explored AI-supported assessment in photography, demonstrating the potential for Convolutional Neural Networks (CNNs) to classify images based on technical quality. Such systems provide valuable feedback for improving technical proficiency but often lack the capability to assess the creative and subjective qualities of an image. Moreover, many existing tools in photography education offer only summative feedback, which does not facilitate iterative improvement and skill refinement, both of which are critical for creative learning. Unlike these existing systems, M-Stock aims to bridge this gap by integrating both technical and creative evaluations, providing formative feedback that encourages continuous learning. The system's feedback is designed not only to assess basic technical aspects but also to guide students in enhancing their artistic interpretation and aesthetic sensibilities, offering a more comprehensive educational experience.

### 2.3. Convolutional Neural Networks for Image Classification

Convolutional Neural Networks (CNNs) have emerged as a robust tool for image classification, widely applied in various fields, including medical imaging, autonomous driving, and creative media [12]. CNNs excel at identifying spatial hierarchies and features in visual data, making them well-suited for assessing technical quality in photography. While CNNs have demonstrated high accuracy in image classification, most studies in this area have focused on technical metrics without exploring how these models might be adapted to assess creative and subjective qualities in educational contexts. Other deep learning models, such as transformers and attention-based networks, have also shown success in visual tasks, providing an alternative to CNNs. However, CNNs remain the primary choice for this study due to their well-established efficiency and proven effectiveness in photography-related tasks. Future iterations of M-Stock could explore alternative models or ensemble approaches to further enhance its evaluative capabilities, particularly for assessing creativity.

### 2.4. Existing Gaps in Automated Photography Assessment

Despite the advances in AI-based assessment tools, significant gaps remain in the automated evaluation of creative student outputs. Most current systems excel at objective assessments, but they struggle to capture subjective elements, such as artistic style

and emotional impact, which are essential in photography education [13]. Additionally, the lack of iterative, formative feedback in current photography assessment tools limits their effectiveness in supporting continuous skill development. The need for systems that can provide nuanced, ongoing feedback on both technical and creative elements of student work remains largely unmet. In response to these challenges, M-Stock was designed to provide a balanced approach to automated photography assessment, incorporating both technical and creative evaluations. By integrating AI-based formative feedback, M-Stock addresses the limitations of existing systems, offering students timely, constructive feedback that promotes self-directed learning and skill enhancement.

## 3. Proposed Method

The M-Stock system was developed to automate the assessment of student photography, providing a balanced evaluation that addresses both technical and creative aspects of students' work. This section outlines the methodology used to design, implement, and evaluate the M-Stock system, focusing on data collection, model training, feedback mechanisms, and system architecture.

### 3.1. Data Gathering

The M-Stock system's training dataset combines images from two primary sources to cover diverse photography skills and quality levels: *Student Assignments from Photography Courses:* Images were collected from photography courses at Mae Fah Luang University. These assignments covered various topics such as fast shutter speed, long shutter speed, night light photography, composition and subject, aperture and depth of field, light and shadow, portrait photography, moving subjects, and product photography. The assignments were submitted via Google Classroom [14], and each image was categorized into three performance levels: Excellence, Good, and Bad, based on criteria established by instructors and photography experts as shown in Figure 2.
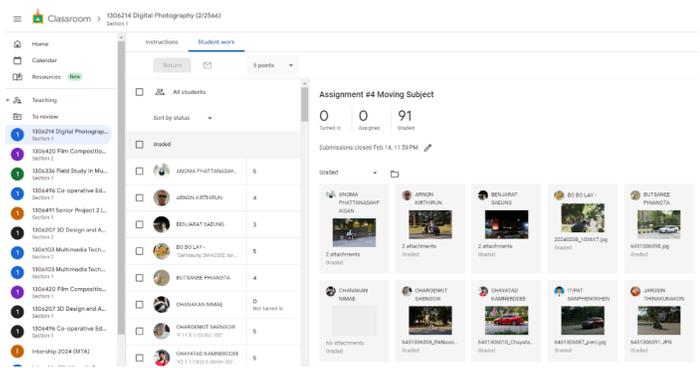


Figure 2: Digital Photography Assignment in Google Classroom

Commercial Standards from Shutterstock Submissions: To integrate professional criteria, the dataset includes student submissions to Shutterstock [15], labelled as either Accepted (commercially viable) or Rejected (commercially inadequate). This source introduces real-world standards into the model, making it robust for assessing quality in a manner that aligns with industry requirements as depicted in Figure 3. Images were stored in a server database, organized by assignment type and quality

category. This collection strategy ensures that the M-Stock model can generalize well across different photography styles, skill levels, and educational contexts.
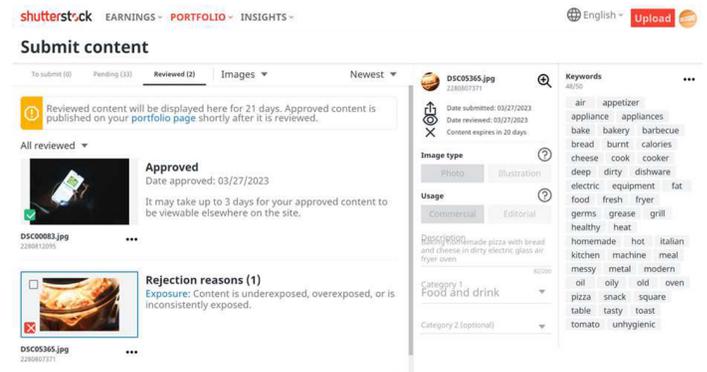


Figure 3: Assessment of uploaded photographs from Shutterstock

### 3.2. Model Training and Selection

The M-Stock system utilizes Convolutional Neural Networks (CNNs) [16] due to their strong performance in visual data analysis and spatial feature extraction. CNNs were chosen over alternative models, such as transformers, because of their efficiency in handling complex image data with lower computational requirements, making them suitable for real-time feedback in educational environments. The CNN architecture includes multiple convolutional layers, ReLU activations, batch normalization, max-pooling layers, and fully connected layers. A final softmax classifier predicts image categories (e.g., Excellence, Good, Bad, Accepted, Rejected), as shown in Figure 4.
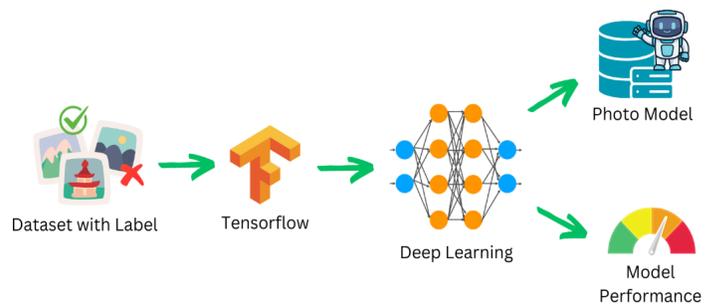


Figure 4: Model training and selection diagram

The training process involved the following steps:

*Image Preprocessing:* All images were resized to 800 x 800 pixels to maintain consistency, ensuring that the model could effectively extract meaningful features across various image types. *Model Optimization:* The Adam optimizer [17] was used to minimize the loss function (sparse categorical cross-entropy), ensuring that the model converged efficiently [18]. During training, performance metrics such as accuracy, prediction speed, and training time were monitored to evaluate the model's effectiveness. During training, performance metrics such as accuracy, prediction speed, and training time were monitored to evaluate the model's effectiveness. The training process was executed using Python 3.11.0 [19], Keras [20], and TensorFlow 2.13 [21]. *Evaluation Metrics:* In addition to accuracy, other metrics such as precision, recall, and F1 score were used to

comprehensively assess the model's effectiveness. These metrics are essential in ensuring that the system's predictions are reliable across different types of assignments and quality levels.

## 3.3. Web Implementation and User Interface

The third component of the M-Stock system is the development of a user-friendly web application that allows students and instructors to interact with the model in real time. The web application was developed using Streamlit 1.31.0 [22], a Python-based framework that simplifies the deployment of machine learning models in web environments. Users initiate the M-Stock system by accessing the website via the URL http://datascience.mfu.ac.th/mstock/. The application's user interface is designed to be intuitive, enabling students to submit their photographs for assessment quickly and easily, Figure 5 illustrates the user interface of the homepage.
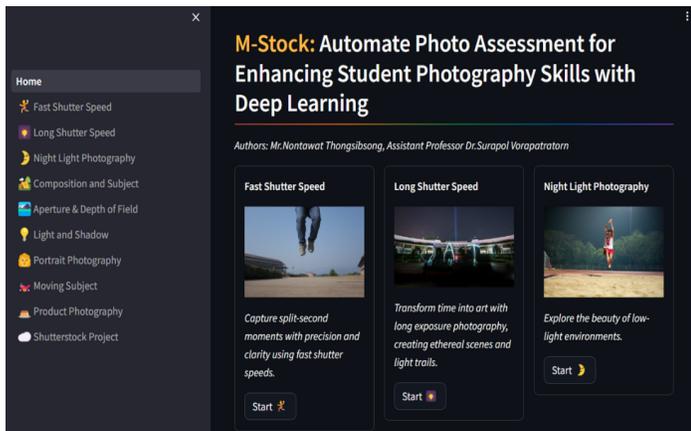


Figure 5: The user interface of the homepage

The submission process involves the following steps: Image Upload: Students select the assignment type and upload their photographs via the web interface. The system supports image formats such as JPG and PNG, with a maximum file size of 200 MB per image. Image Preprocessing and Classification: Once an image is uploaded, the system preprocesses it by resizing and standardizing the input. The pre-trained CNN model then classifies the image, providing a prediction and confidence score for each category (e.g., Excellence, Good, Bad). Feedback Delivery: The classification results are displayed immediately, allowing students to receive prompt feedback on their work. This feedback can help students identify areas for improvement and refine their photography skills iteratively. The M-Stock web application includes 11 pages: one homepage and ten assignment pages. Each assignment page corresponds to a specific photography lesson, where students can view sample photographs and submit their own work for evaluation. The web application's architecture ensures that it can scale to accommodate larger datasets and more complex assignments as the photography curriculum evolves. The Assignments page's user interface is depicted in Figure 6.

Overall, the M-Stock system combines the power of CNN-based image classification with a tailored feedback mechanism to support student learning in photography. Through its combination of technical rigor, creative assessment, and real-time feedback, M-Stock offers a novel solution for enhancing photography education in university settings. This method ensures that students receive immediate, meaningful feedback on their work, fostering continuous improvement and skill development in both technical and artistic aspects of photography.
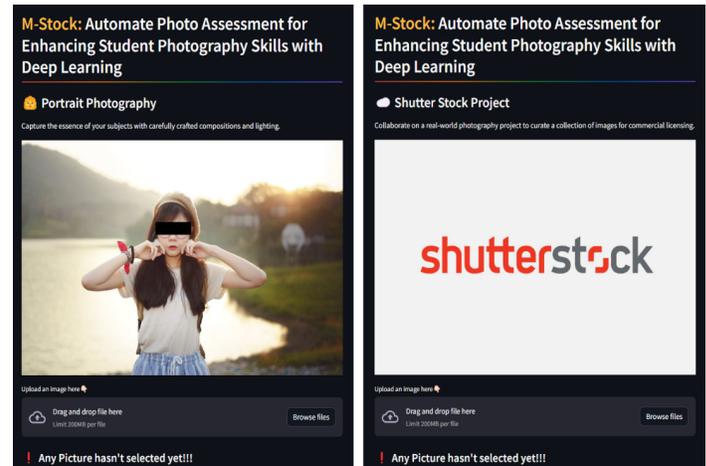


Figure 6: The user interface of the assignment page

## 4. Evaluation And Results

The M-Stock system was evaluated based on its classification accuracy, prediction speed, and training time, along with additional metrics such as precision, recall, and F1 score to provide a comprehensive assessment. Furthermore, a pilot study was conducted with students to gather qualitative feedback on their learning experience with M-Stock. This section presents the experimental setup, results, and analysis, demonstrating M-Stock's efficacy in supporting photography education.

### 4.1. Experimental Setup

The M-Stock system was tested in a virtualized server environment using VMware ESXi [23], running Windows Server 2016 [24] with an 8-core CPU (2.10 GHz) and 16 GB of RAM. The dataset for training and testing included 4,616 student images from various photography assignments and 244 Shutterstock images. and the necessary software tools, including Python 3.11.0, Keras with TensorFlow 2.13.0, and Streamlit 1.31.0 for web deployment. The dataset was divided into an 80% training set and a 20% test set, ensuring a robust model capable of handling diverse image categories. The system's scalability and performance were also evaluated under different image resolutions and batch sizes. Additionally, a small-scale pilot study with 30 students was conducted to assess the impact of M-Stock's feedback on learning outcomes.

### 4.2. Model Performance

The M-Stock system was evaluated for its ability to accurately classify student photography submissions across various assignment types, including technical and creative tasks. To assess the model's effectiveness, we measured several key metrics: accuracy, precision, recall, and F1 score for each assignment type. These metrics provide a comprehensive view of the model's classification performance, highlighting its strengths in technical precision and adaptability to different photography genres. The results are presented in Table 1

Table 1: Performance of each photo model for the M-Stock System

| Photo Model | Accuracy (%) | Precision | Recall | F1-Score | Training Time (min.) | Prediction Speed (ms) |
|---|---|---|---|---|---|---|
| Fast Shutter Speed | 96.72 | 0.96 | 0.95 | 0.96 | 28.7 | 47.1 |
| Long Shutter Speed | 96.93 | 0.97 | 0.96 | 0.97 | 39.1 | 46.7 |
| Night Light Photography | 98.36 | 0.98 | 0.97 | 0.98 | 45.5 | 50.6 |
| Composition and Subject | 98.77 | 0.99 | 0.98 | 0.99 | 70.4 | 43.7 |
| Aperture, Depth of Field | 95.33 | 0.95 | 0.94 | 0.95 | 14.7 | 47.9 |
| Light and Shadow | 95.47 | 0.94 | 0.95 | 0.94 | 35.4 | 45.3 |
| Portrait Photography | 99.53 | 0.99 | 0.99 | 0.99 | 119.2 | 43.4 |
| Moving Subject | 97.54 | 0.97 | 0.96 | 0.96 | 16.6 | 46.5 |
| Product Photography | 96.73 | 0.96 | 0.95 | 0.96 | 14.3 | 44.7 |
| Shutterstock Project | 96.39 | 0.95 | 0.94 | 0.94 | 20.3 | 45.2 |
| **Total (Average)** | **97.18** | **0.97** | **0.96** | **0.96** | **40.4** | **46.1** |

*Accuracy:* The system achieved an overall accuracy of 97.18%, with individual assignment accuracies ranging from 95.33% for Aperture, Depth of Field to 99.53% for Portrait Photography. The high accuracy demonstrates M-Stock's ability to consistently classify images across diverse photography tasks, from technical skills (e.g., Long Shutter Speed) to more composition-focused assignments (e.g., Composition and Subject). High accuracy in these varied tasks indicates that M-Stock can generalize well across different photographic techniques and styles, making it adaptable to a comprehensive photography curriculum. In addition to accuracy, we calculated precision, recall, and F1 scores for each assignment type to gain insights into M-Stock's classification reliability: *Precision:* High precision values (average of 0.97) indicate that M-Stock has a low rate of false positives, meaning it rarely misclassifies lower-quality images as higher quality. This is crucial in an educational context where students need accurate feedback to understand areas requiring improvement. *Recall:* The average recall of 0.96 shows M-Stock's effectiveness in identifying all images that meet specific quality criteria. High recall is especially important for technical assignments, as it ensures that the system accurately identifies images with correct exposure, composition, and other technical parameters. *F1 Score:* With an average F1 score of 0.97, M-Stock demonstrates a balanced performance in both identifying correct classifications and avoiding misclassifications. This score, the harmonic mean of precision and recall, confirms that the system provides reliable feedback, balancing sensitivity and specificity. The average prediction speed of 46.1 milliseconds per image shows that M-Stock provides rapid feedback, which is essential in real-time educational environments where students submit images and expect prompt responses. This quick feedback loop enables students to immediately identify mistakes and make improvements, reinforcing the learning process. The system's training time varies based on assignment type, with more complex tasks such as Portrait Photography taking longer (119.2 minutes) due to the intricate analysis required.

M-Stock's classification performance metrics demonstrate its effectiveness in providing real-time feedback across a wide range of photographic techniques. By maintaining high accuracy, precision, and recall across both technical and creative assignments, the system supports educators in delivering consistent, objective feedback to students. This capability is particularly beneficial in large classes, where individualized feedback is challenging to provide manually. With M-Stock, students can receive accurate, actionable feedback that promotes self-directed learning and skill refinement. Overall, M-Stock's classification performance confirms its suitability as a comprehensive educational tool, capable of assessing diverse photography tasks with high accuracy and efficiency. Future enhancements may involve refining these classification models further to increase precision and recall in more subjective creative categories, aligning with the evolving needs of photography education.

### 4.3. Scalability and Runtime Performance

The scalability of the M-Stock system was tested across various image resolutions and batch sizes to evaluate its capacity for handling large volumes of submissions in real-time classroom settings. Scalability is essential in educational applications where a high number of images may be submitted simultaneously, especially in large classes. The system was assessed under four different image resolutions—640x480 (low), 1280x720 (HD), 1920x1080 (Full HD), and 3840x2160 (4K)—to analyse the effect of image size on prediction speed and accuracy. For each image resolution, we measured the average prediction speed, batch processing time, and accuracy to determine the system's efficiency and robustness under increasing data sizes. Table 2 below illustrates these findings:

Table 2: Different Image Sizes Experiment Results

| Image Size | Prediction Speed (ms) | Processing Time (Sec) | Acc. (%) |
|---|---|---|---|
| 640 x 480 | 39.5 | 2.1 | 96.3 |
| 1280 x 720 | 46.1 | 2.5 | 97.2 |
| 1920 x 1080 | 52.4 | 3.2 | 97.8 |
| 3840 x 2160 | 74.6 | 5.4 | 98.1 |

These results show that the system maintains high accuracy across all resolutions, with a minimal decrease in prediction speed as image size increases. For low and HD resolutions, prediction times are under 50 milliseconds, allowing near-instantaneous feedback in real-time applications. Full HD and 4K images take slightly longer to process, but the prediction speeds are still well within acceptable limits for classroom use, ensuring efficient

operation even for high-quality images. The accuracy remains high across resolutions, demonstrating that M-Stock's performance does not degrade with larger image sizes. The system's batch processing ability was evaluated to simulate high-demand situations where multiple students submit images simultaneously. We processed batches of 50 images at different resolutions, recording the total processing time required. M-Stock handled batch submissions with only a slight increase in processing time for higher-resolution images, completing a 50-image batch in approximately 2.1 seconds at low resolution and 5.4 seconds at 4K. This capability indicates that M-Stock is well-suited to handle real-time feedback needs in large classes, where simultaneous submissions are common.

## 4.4. User Satisfaction

To assess user satisfaction with the M-Stock system, a survey was conducted among 30 students and 5 instructors during the pilot study. The survey evaluated four key dimensions: ease of use, feedback clarity, perceived usefulness, and overall experience. Participants rated each dimension on a Likert scale from 1 (strongly disagree) to 5 (strongly agree). The results are summarized in Table 3.

Table 3: User Satisfaction Survey Results

| Dimension | Average | SD | Agreement (Score ≥ 4) |
|---|---|---|---|
| Ease of Use | 4.7 | 0.3 | 93% |
| Feedback Clarity | 4.5 | 0.4 | 87% |
| Perceived Usefulness | 4.6 | 0.5 | 90% |
| Overall Experience | 4.6 | 0.3 | 92% |

The survey results indicate high levels of satisfaction across all dimensions. Students found the system's interface intuitive and straightforward, with an average score of 4.7 for ease of use. Feedback clarity received an average score of 4.5, reflecting the comprehensibility of the AI-generated evaluations. The system's ability to enhance photography skills was rated 4.6 on average, indicating its perceived effectiveness in promoting self-directed learning. Overall, users rated their experience with the system highly, with an average score of 4.6 and 92% agreement. Qualitative responses also highlighted specific benefits, such as the speed of feedback delivery and the ability to focus on iterative improvement. Some suggestions for enhancement included adding more nuanced assessments of creative aspects, such as artistic style and emotional impact. The results demonstrate that M-Stock effectively supports both teaching and learning objectives, providing a user-friendly, impactful solution for photography education.

Students also reported appreciating the quick turnaround time of feedback, which allowed them to adjust in near real-time. These findings suggest that M-Stock's formative feedback supports continuous learning, enhancing students' technical and creative skills. The results show that the M-Stock system performed exceptionally well in both educational and commercial contexts. The high accuracy rates across all categories demonstrate that the CNN model is capable of handling diverse photographic styles and quality levels. The relatively low prediction speed of 46.1 milliseconds per image allows the system to provide immediate

feedback, which is crucial for enhancing the learning experience in photography courses. The results of the photo quality assessments for each assignment are displayed in Figure 7.
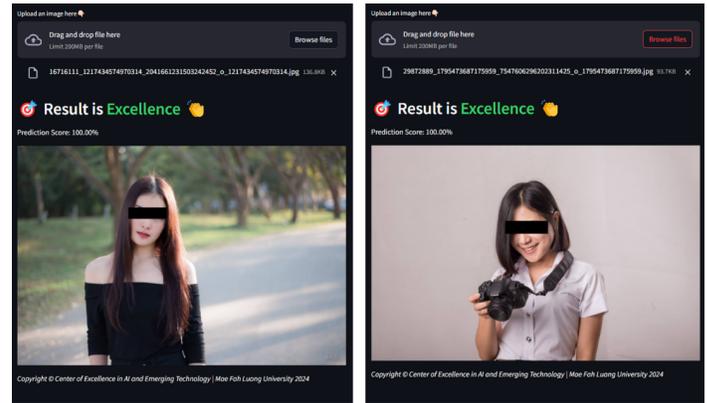


Figure 7: The photo quality assessment in 'Excellence' result

The portrait photography model, which achieved the highest accuracy (99.53%), required the longest training time (119.2 minutes). This indicates that more complex assignments, which involve intricate features such as lighting and composition in portrait photography, require more computational resources to train effectively. However, once trained, the model can classify images quickly and accurately. In contrast, simpler assignments, such as Product Photography and Moving Subject, required significantly less training time but still achieved high accuracy, indicating that the model can generalize well across different photography styles. The Shutterstock project data also yielded strong results, with an accuracy of 96.39%. This indicates that the system can meet industry standards for evaluating commercial photography, providing feedback that aligns with professional evaluation criteria. The results of the photo quality assessments for the Shutterstock project are presented in Figure 8.
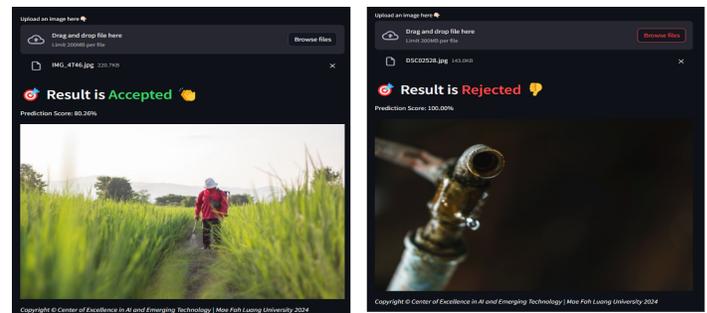


Figure 8: The photo quality assessment in 'Excellence' result

The evaluation results indicate that M-Stock performs reliably across technical metrics, while the pilot study confirms its positive impact on student learning. The high accuracy, coupled with quick feedback delivery, underscores M-Stock's suitability for real-time educational applications. The scalability tests further demonstrate that the system is robust enough to handle diverse classroom environments with high submission volumes. Overall, M-Stock provides a comprehensive assessment experience for photography students, offering both technical precision and creative guidance. Future work may explore expanding the system's feedback capabilities to include more nuanced assessments of creative

elements, potentially incorporating reinforcement learning techniques to adapt feedback based on individual student progress.

## 5. Conclusion

The M-Stock system represents a significant advancement in photography education by leveraging the power of artificial intelligence to provide automated, real-time feedback on both technical and creative aspects of student submissions. By utilizing Convolutional Neural Networks (CNNs), the system achieved high accuracy (97.18%) and rapid prediction speeds (46.1 milliseconds per image), making it a reliable and scalable solution for dynamic classroom environments. Through a combination of quantitative evaluations and qualitative user feedback, the study demonstrated that M-Stock effectively enhances student learning experiences. Students reported improvements in their technical skills, self-directed learning, and overall engagement, while instructors appreciated the system's ability to maintain consistent evaluation standards across large class sizes. The system's ease of use and comprehensive feedback mechanisms make it a valuable tool for fostering continuous learning and skill development in photography courses.

Despite these achievements, challenges remain in assessing highly subjective creative elements, such as artistic style and emotional impact. Future iterations of M-Stock should incorporate advanced techniques, such as reinforcement learning or generative models, to provide deeper insights into these aspects. Additionally, expanding the platform to support other creative disciplines, such as graphic design and visual arts, could broaden its applicability and impact. In conclusion, M-Stock exemplifies how AI can transform education by addressing key limitations of traditional assessment methods. By combining technical rigor with creative evaluation, the system not only meets the evolving needs of photography education but also sets the stage for broader applications of AI in creative and technical learning environments.

## Conflict of Interest

The authors declare no conflict of interest.

## Acknowledgment

## References

[1]  T. Crawford, C. DeLaney, Starting your career as a freelance photographer, Simon and Schuster, 2017.

[2]  H. Alkan, O. Topuz, B. İnce, Ş. Kapıkıran, "The effects of basic photography education on quality of life, self-esteem, life satisfaction and moods in children with diplegic cerebral palsy: A randomized controlled study," Physical & Occupational Therapy in Pediatrics, **42**(1), 1–11, 2021.

[3]  C.-M. Chen, "Personalized E-learning system with self-regulated learning assisted mechanisms for promoting learning performance," Expert Systems with Applications, **36**(5), 8816–8829, 2009.

[4]  T. Zoltie, T. Shemwood, "Instructional design of a clinical photography course for undergraduate dental students," Journal of Visual Communication in Medicine, **42**(2), 47–51, 2019.

[5]  M. Munakata, A. Vaidya, "Encouraging creativity in mathematics and science through photography," Teaching Mathematics and Its Applications: International Journal of the IMA, **31**(3), 121–132, 2012.

[6]  Z. Yanhua, L. Jiaogang, "Research on the expression of photography language under digital technology," in 2020 5th International Conference on Electromechanical Control Technology and Transportation (ICECTT), IEEE: 228–230, 2020.

[7]  M.C. Sahin, "Instructional design principles for 21st century learning skills," Procedia-Social and Behavioral Sciences, **1**(1), 1464–1468, 2009.

[8]  D. Freer, Microstock photography: how to make money from your digital images, Routledge, 2008.

[9]  K. Kemavuthanon, "Integrated E-project collaborative management system: Empirical study for problem-based learning project," in 2017 9th International Conference on Information Technology and Electrical Engineering (ICITEE), IEEE: 1–5, 2017.

[10]  L. Jacobs, Professional commercial photography: techniques and images from master digital photographers, Amherst Media, 2010.

[11]  N. Thongsibsong, K. Kemavuthanon, S. Vorapatratorn, "Enhancing Student Photography Skills with Web-based Photo Stock and Learning System," in 2023 20th International Joint Conference on Computer Science and Software Engineering (JCSSE), IEEE: 454–457, 2023.

[12]  A. Krizhevsky, I. Sutskever, G.E. Hinton, "ImageNet classification with deep convolutional neural networks," Communications of the ACM, **60**(6), 84–90, 2017.

[13]  N. Thongsibsong, "Photography Supporting System for Distance Learning in University," in 2021 13th International Conference on Information Technology and Electrical Engineering (ICITEE), IEEE: 15–18, 2021.

[14]  B. Sahu, "Digital Tools for Educational Enhancement," Digital Narratives in Education, 78, 2024.

[15]  Q. Jadwan, "How to get Money From Shutterstock Contributor in Smartphone," Hikamatzu| Journal of Multidisciplinary, **1**(1), 7–12, 2024.

[16]  T. Kattenborn, J. Leitloff, F. Schiefer, S. Hinz, "Review on Convolutional Neural Networks (CNN) in vegetation remote sensing," ISPRS Journal of Photogrammetry and Remote Sensing, 173, 24–49, 2021.

[17]  A. Keras, "Keras 3 API documentation," Metrics, Regression Metrics Https://Keras. Io/Api/Metrics/Regression_metrics, 2024.

[18]  M. Yeung, E. Sala, C.-B. Schönlieb, L. Rundo, "Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation," Computerized Medical Imaging and Graphics, 95, 102026, 2022.

[19]  G. VanRossum, F.L. Drake, The python language reference, Python Software Foundation Amsterdam, The Netherlands, 2010.

[20]  E. Dumić, "Learning neural network design with TensorFlow and Keras," in ICERI2024 Proceedings, IATED: 10689–10696, 2024.

[21]  T. Developers, "TensorFlow," Zenodo, 2022.

[22]  A. Streamlit, "Streamlit: A faster way to build and share data apps," Faster Way to Build and Share Data Apps, 2024.

[23]  L. Patrão, VMware and vSphere Overview, Springer: 9–18, 2024.

[24]  J. Krause, Mastering Windows Server 2016, Packt Publishing Ltd, 2016.

# A Review of Natural Language Processing Techniques in Under-Resourced Languages

Stephen Obare[*], Kennedy Ogada

*Jomo Kenyatta University of Agriculture and Technology, ICSIT, Nairobi, Kenya*

A R T I C L E   I N F O

A B S T R A C T

*Natural language processing (NLP) techniques have transformed a number of tasks in the modern age of information explosion where millions of gigabytes of data are generated every day. Despite achieving state-of-the-art performance in high-resource languages, current techniques struggle with processing under-resourced languages which are characterized by data scarcity, linguistic diversity, computational limitations, ambiguity of language syntax and semantics. In this paper, we first introduce domains of application of NLP techniques, the limitations of current approaches and our contribution. Second, we investigate the progress and challenges that hinder NLP techniques from being equitable and useful to under-resourced languages. We then discuss opportunities for developing more inclusive NLP techniques that allow everybody, everywhere – rich or poor – to access all the advantages of advanced language technologies and at the same time preserve global linguistic diversity.*

## 1. Introduction

Natural language processing (NLP) stands halfway between computer science computational linguistics, and it is dedicated to the conversion of written and spoken natural human languages into structured mineable data. Through the combination of linguistic, statistical and artificial intelligence (AI) methods NLP can be used either to determine the meaning of a text or even to produce a human-like response. In recent decades, rapid development in NLP has resulted in great performance breakthroughs in various domains including healthcare, education, finance, e-commerce, etc. Although NLP techniques have advanced from rule based methods to deep learning based transformer models, NLP has opened new frontiers to understand, generate, or even translate human language more accurately and more efficiently [1].

NLP is already part of our everyday life as it is widely implemented in our computer software or in our mobile phones [2]. Other areas of application include medical diagnostics as well as personal healthcare [3]. For example, a recent work showed how AI accelerated multiomics integration to enable predictive modeling of disease [4]. Other NLP-powered question answering systems have also been developed for improving medical research including in hepatocellular carcinoma [5]. The applications provided by these applications demonstrate uses of NLP to obtain actionable insights from unstructured clinical data.

NLP and AI integration is the new trend in the education sector, making content delivery and accessibility easy and improved [6]. In [7], the authors used NLP to study its role in text summarization, sentiment analysis, and domain-specific keyword extraction. For instance, transformer-based models such as Baichuan2 Sum have been trained for dialogue summarization and have improved interaction in the educational settings [8]. NLP techniques have also taken advantage of the financial domain, specifically in regards to analysing consumer reviews and predicting market trends [9]. Other authors [10] deployed NLP to refine consumer insights by attribute embedding to create hierarchical representation of product reviews. In addition, NLP is employed to boost the cybersecurity resilience of financial systems by capturing the nature of fraudulent patterns and thwarting threats [11].

Yet, NLP still has a long way to go. One of the challenges experienced by current techniques when processing under-resourced languages is the data scarcity problem which is a small dearth of large high quality datasets [1]. Advanced techniques based on deep learning models, also referred to as 'black boxes', are nonetheless uninterpretable, posing a danger in sensitive applications [12]. Explainable AI (XAI) and retrieval augmented text generation [13] have been proposed as techniques towards ethical deployment of NLP systems. This paper explores the limitations of current techniques when processing under-resourced languages and presents opportunities for developing more inclusive NLP techniques that

---

[*]Corresponding Author: Stephen Obare, 47210, 00100, smobareo@gmail.com

allow everybody, everywhere – rich or poor – to access all the advantages of advanced language technologies and at the same time preserve global linguistic diversity.

We make the following contributions:

1. A review of NLP processing techniques for under-resourced languages is presented.

2. We probe challenges and open problems experienced by state-of-the are techniques.

3. We highlight main challenges and future research directions for processing under-resourced languages.

The rest of the paper is organized as follows. We present the cause of the problems, challenges and related work in Section 2. Section 3 outlines some of the approaches that have been developed to address the challenges, their limitations and proposed solutions. Section 4 provides an analysis of existing models, lingering gaps and several promising directions for making NLP more usable for under-resourced languages. Finally, conclusions and future work are presented in Section 5.

## 2. Background

In recent years NLP has made significant progress due to the availability of large scale data and computational power. However, these breakthroughs have disproportionately helped languages with rich linguistic resources which are referred to as well-resourced languages [14]. On the other hand, NLP suffers under serious a shortage of resources for under resourced languages, and this is why a large portion of the world's languages are challenged in NLP. What causes these challenges lies in the fact that such data are scarce, they are difficult to handle, and the existing NLP systems suffer from biases. Solving these challenges is at the core of powering NLP technologies for a broader range of languages, the use of which spans marginalized communities [15]. This section expounds on some of these challenges.

### 2.1. Data scarcity

Scarcity of large annotated corpora is one of the main issues in acquiring NLP tools for under-resourced languages. Ranathunga et al., 2022 [14] established that many NLP models require large amounts of annotated data to learn language patterns, semantic relationships, and syntactic structures. Such corpora are readily available for well-resourced languages such as English, Spanish and Chinese, through public datasets, digital content, and linguistic resources.

However, the data needed for training high performance models are not available for the under-resourced languages such as Sheng which are mostly found in developing countries [7]. Furthermore, in many of these communities, the under-resourced languages are spoken and the technology infrastructure necessary to generate digital content is not in place [16]. The result of this is that these languages are not well represented with the limited data available in digital platforms.

### 2.2. Language complexity

High levels of morphological richness, syntactic diversity are characteristic of under-resourced languages, rendering them difficult to represent in NLP systems. As an example, agglutinative languages like Turkish or Finnish requires models to deal with intricate morphological structures that have built up large amounts of grammatical encoding in the form of prefixes and suffixes [4]. Like Sheng, especially ergative-absolutive languages, they present syntactic puzzles that are not present in well-resourced languages.

### 2.3. Multiplicity of languages

Added to the complexity described in Section 2.2 is the fact that many under-resourced languages are in multiple dialects or oral traditions with no formal orthography [12]. Tokenization, lemmatization, and other preprocessing tasks become that much more difficult in the face of these factors when applied to under-resourced languages.

### 2.4. Bias

Another important challenge with NLP systems in under-resourced languages is bias. The ability of models trained on datasets for linguistically diverse languages to generalize to languages where resources are scarce has been shown to fail by providing erroneous translations or sentiment analyses [11]. For example, cultural nuances as well as idiomatic expressions that are native to under-resourced languages are often misrepresented or completely ignored [5].

Such biases create digital inequalities further while they lead to inaccuracies. For instance, machine translation systems can output suboptimal outputs for the under-resourced languages, widening the digital divide in which speakers of such languages will be excluded from benefits of the NLP driven technologies [1]. As a way to add another way to marginalize communities, this will also allow the erosion of linguistic diversity in the digital space.

## 3. Existing approaches

To address these gaps identified in Section 2, innovative strategies, including, for example, transfer learning, multilingual pretraining, and architectures designed for low resource languages are needed. Further democratization of NLP for under-resourced languages can be achieved through collaborative efforts in creating open source datasets, crowd-sourced linguistic resources or efficient fine tuning methods. We present some of the techniques in this section.

### 3.1. Wordnets

WordNets are a core resource and fundamental enabler for machines to acquire word meaning and meaning relationship [17]. Wordnet is a lexical database that groups words in sets of synonyms, or synsets, along with semantic relationships between these sets, like hyponymy (a general term and a less general one) meronymy (one part of another), and antonymy (opposites). To some extent,

Wordnets became an ontology understood as a collection of semantically reduced hierarchies and associations necessary for several NLP tasks [18].

WordNets are at their core structured representations of the linguistic knowledge, scalable meaning of words, and their relationship to each other. The English WordNet is the most widely known and used WordNet which was instrumental for a variety of NLP tasks like word sense disambiguation (WSD), semantic parsing, and machine translation [1] [19]. Applications, including question answering and search, automatic summarization, need WordNets these map words to synsets and provide semantic relation that make language processing easier [1].

It has been challenging but promising effort to extend WordNets to other languages, especially to cover under-resourced languages. With projects such as the Universal Wordnet, the goal is to create multilingual Wordnets so as to facilitate cross-lingual NLP tasks, e.g. translation alignment, and semantic equivalence mapping [20]. This development provides the possibility of having Wordnets that facilitate linguistic diversity in NLP.

Through community driven initiatives, Global Wordnet Association [21] has acted as an important force behind Wordnet development for under-resourced languages. For example, the Indian Wordnet project has developed lexical databases for lesser used regional Indian languages for the country's linguistically diverse population [22].

### 3.1.1. Limitations

While Wordnets are important for processing under-resourced languages, Wordnets have limitations. A major problem is lack of domain specific vocabulary [23]. Wordnets are able to cover the generality most comprehensively, however, specialized entries for field-specific terms (medicine, engineering, or law) are typically absent [24]. The performance of NLP systems in domain-specific tasks, i.e., clinical text analysis, or legal document review, is impaired by this deficiency [25].

Furthermore, building Wordnets for morphologically complex languages such as Sheng, Finnish or Turkish continues to be a hard problem. With rich inflectional morphology, and with a large number of word forms derived from a single root, Wordnets for these languages require extensive linguistic resources and expertise [26]. This makes matters even more difficult when orthography, or even dialectal variations complicate the task, particularly in languages whose traditions are oral or, if written, are written by several scripts [27].

One other limitation of Wordnets is multilinguality. Although the Universal Wordnet is an effort to unify Wordnets across languages, many languages are under-resourced, and thus have no corresponding resources. However, when multilingual Wordnets are present, they are likely not to be perfectly aligned and this could limit its efficiency in crosslingual NLP applications [28].

### 3.1.2. Future direction

Given the semantic nature of many NLP tasks, Wordnets continue to be one of the most valuable tools for furthering NLP, both in

problems of semantic parsing, word sense disambiguation, and multilingual translation. But, gaps in *coverage*, *domain specific vocabulary*, and *linguistic complexities* [29] prevent their development for under resource languages. The gaps will need to be addressed with a focus on creating a complete, multi-lingual, and domain specific WordNets to enable the development of more inclusive and effective NLP applications.

### 3.2. Contextualized models

With contextualized word representations such as Bidirectional Encoder Representations from Transformers (BERT) and Embeddings from Language Models (ELMo), NLP has been revolutionized to capture word meaning dynamically based on their context of use. In contrast to static word embeddings such as Word2vec or GloVe where the same vector is attributed to each word, regardless of the context, the contextualized models modify word representations depending on their context [18]. Thus they are able to cope with polysemy, for example, 'bank' as a financial institution, a riverbank, 'bank' as a verb thereby greatly improving performance in tasks ranging from machine translation, sentiment analysis, named entity recognition, text generation, and document classification [1].

The models are trained over large scale language corpora to pick the word probability within a defined context. For example, BERT with a bidirectional transformer architecture, ELMo with a long short term memory (LSTM) bidirectional network [30].

### 3.2.1. Limitations

Contextualized models have been shown to perform better than traditional techniques in disambiguating terms and inflected forms found in morphologically rich languages [31]. Although contextualized models are successful in well resourced languages, they suffer from the problem of *data scarcity*. These models require massive pre-training on massive corpora of annotated data, which are generously available in well-resource languages but lacking in under-resourced languages [5]. Due to this limitation, contextualized models trained on multilingual datasets (e.g., mBERT or XLM-R) under represent under-resourced languages, and so perform poorly in a number of NLP tasks including named entity recognition and machine translation.

Furthermore, under-resourced languages are also characterized by language diversity. Many of these languages are highly morphologically complex such that a single word can be used in many forms according to the grammatical rules of the languages concerned. For example, Sheng and Turkish languages are agglutinative and have long inflected words; BERT and ELMo have problems generalizing well [32]. In the absence of formalized orthographies or standard spellings in under-resourced languages, the noisy and inconsistent data pose processing challenges to the contextualized models.

Like other contextualized models, training models for under-resourced languages consumes a lot of computational resources and high end GPUs or TPUs are not readily available to most researchers working on these languages. Moreover, the success of multilingual models is dependent on high resources languages dominating the training corpus of multilingual models, which limits their ability

---

[1]https://wordnet.princeton.edu/publications

to represent linguistic features of under-resourced languages [2] [33]. This is because low resource languages in multilingual models have sparse, low quality representations, and thus biased resulting in incomplete outputs on NLP tasks such as sentiment analysis or text classification.

A third critical issue is the quality of the textual data used for training. Since digital content for under-resourced languages is often informal, with non-standard spellings, or too poor in linguistic richness to train high performing models [34]. As a result, it is difficult for contextualized models to learn precise syntax and semantics in such smaller or under-resourced language communities [20].

### 3.2.2. Future direction

Future work could explore utilizing techniques such as transfer learning, cross-linguistic modeling and low resource training strategies. Lightweight models optimized for resource-limited settings could also make NLP advancements more widely available to under-resourced languages democratizing access to them and increasing the field's inclusivity. As more innovation and more collaboration continue, these contextualized models can become more inclusive, and can bridge the gap between well-resourced and under-resourced languages thereby unleashing the promise of those languages representing the linguistic diversity of the world.

### 3.3. Pre-trained language models (PLMs)

Modern NLP has taken a trajectory, wherein PLMs, including GPT (Generative Pre-trained Transformer), are a crucial part of the landscape. PLMs, like their predecessors, make use of a great deal of text corpora to predict masked words or the next word in a sequence with unsupervised learning, capturing syntactic, semantic, and contextual information [35]. In many applications such as sentiment analysis, machine translation, named entity recognition, and text summarization, this ability to generalize across diverse NLP tasks with minimal fine tuning has proven enormously valuable. PLMs have also shown excellent performance for high resource languages such as English and Chinese and have set new benchmarks, allowing the removal of extensive task specific feature engineering [36]. BERT and GPT are trained on billions of tokens, and thus are able to capture complex language pattern and perform intricate tasks at a mindblowing accuracy [1].

### 3.3.1. Limitations

Two major challenges inherent in the PLMs limit their effectiveness in under-resourced languages. The lack of large scale annotated datasets required for effective pretraining is one primary issue. For under-resourced and under-represented languages, the digital presence to build extensive corpora is missing such that PLMs cannot capture their diversity [33].

The morphological and syntactic complexity of many under-resourced languages makes them very challenging for PLMs. There is great morphological complexity in agglutinative languages like Sheng and polysynthetic languages like Inuktitut, making it difficult for PLMs to generalize with great magnitude [37]. Added to that,

there exists no formalized orthographies and inconsistent textual data, rendering it difficult to align language sounds to continuous pitch patterns [38].

Additionally, due to the large proportion of high resource languages in the training dataset, biases are introduced into the features of these under resourced languages which lack idiomatic expressions and grammatical structures [39]. Pretraining and fine tuning PLMs are also computationally resource intensive which prevents their use with under-resourced languages. But training models such as BERT or GPT from scratch requires enormous computational power which is not available to researchers in resource constrained environment. When textual data is available, however, while it could be of sufficient quality, its diversity or even its very existence can render PLMs unable to effectively represent it, with informal content and non-standard spellings entering to influence its accuracy [40].

### 3.3.2. Future direction

To overcome these limitations, there is need for cross-lingual transfer learning, multilingual alignment, and low resource pretraining techniques. Likewise, improving applicability of PLMs to under-resourced languages involves creating open source linguistic resources and lightweight models that are sensitive to resource constraints. Further work with continued innovation and collaboration has the potential to make PLMs more inclusive for representing and processing linguistic diversity of the world [41].

### 3.4. Large language models (LLMs)

LLMs are trained on massive datasets from books, articles, web content and can generate coherent, contextually relevant text; language translation; summary; answering complex questions [40]. But what enables these models to be so large on the scale of billions or trillions of parameters is exactly their ability to model nuanced language patterns, generate poetry or fluent descriptions for an unimaginable number of images, or to emit contextually aware text. Generative NLP with LLMs such as GPT-3 has established new standards, ranging from creative writing to technical application, writing essays, programming code production and conversational simulating. Overall LLMs ability to generalize across many diverse tasks is attributed in part to the fact that they have been pre-trained on massive and diverse datasets [42].

### 3.4.1. Limitations

However, LLMs are powerful but have some issues to overcome when used in under-resourced languages. The scarcity of training data is probably the main limitation. This is especially true since these models need enormous datasets for pre-training, but under resourced languages by definition lack enough digital representation for building such datasets [43]. Multilingual models like GPT-3 frequently exhibit biases towards high resource languages leaving under-resourced languages poorly represented, with poor performance. For these languages, existing tasks such as machine

---

translation, sentiment analysis and question answering are hampered by this imbalance [44].

In addition, LLMs are ill-equipped to address particularly difficult linguistic challenges posed by under-resourced languages. A large number of these languages have morphologically complex structures, deviant syntax, or have no established written form. For a specific example, African and Indigenous languages are typically primarily oral and have insufficient written representation to form LLM training data.

Training LLM models require extensive hardware infrastructure which are typically not only unavailable, but disproportionately costly in disenfranchised language regions. This limitation keeps researchers and developers in these areas from training or adapting LLMs for their particular linguistic needs. When some training data is available, its quality is very low because it include informal content, nonstandard spelling, and sparse domain diversity that severely degrades LLMs' ability to encode it accurately [45].

### 3.4.2. Future direction

Besides the strategies in Section 3.3.2, there is need for concerted effort towards creating well developed datasets for under-resourced languages [46]. More importantly, lightweight models optimized for resource constrained environments may make LLMs more inclusive and accessible. These efforts, with consistent research and collaboration, can allow LLMs to represent the linguistic diversity of the world more effectively, at the intersection between high-resource and under-resource languages [47].

However, as pointed out by techniques such as transfer learning, which improve by adapting beforehand knowledge from high to low resource languages, the underlying bias in pre-trained data most frequently limits how much they help and distort performance, even after fine tuning [48].

## 4. Towards a more inclusive and accessible NLP models

Section 2 and Section 3 summarize the constraints of existing models and techniques due to a number factors including data scarcity, linguistic diversity and computational resources. We attempt in this section to provide a unified view on how do develop models for under-resourced languages in the face of challenges and limitations highlighted above. Whereas well-resourced languages boast rich annotated datasets that enable effective training of high performance NLP models, under-resourced languages do not. Its scarcity makes it difficult to develop reliable models that account for the particularism in these languages' features, such as their dense morphological structures, dense syntax, and distinctive phonology. This deep dictionary of language, combined with the linguistic complexities, makes these limitations acute in light of conventional models which are typically trained on more simple language structures [49].

Another major issue is bias in NLP models, predominantly trained on high resource languages. Often models such as Word-Nets, contextualized models, and large pre trained language models (PLMs)) fail to capture the specifics of under-resourced languages. WordNets offer useful semantic relationships for word sense disambiguation, and the presence of semantics is often desirable for

language pairs with difficult morphological rules, but they often lack domain specific coverage and vocabulary. Like contextualized models, BERT and its ilk, PLMs too, struggle with under-resourced languages because of lack of training data, and while PLMs have advanced capabilities, they are limited by the biases in the datasets on which they were pretrained. Even the most recent progress in NLP, large language models like GPT-3, face as big of a challenge. These models depend on large data sets and high computational requirements, which makes them infeasible for many under-resourced language communities, leading to their inability to perform well on these languages [5].

Though progress has not been without comparison, the analysis of existing models points out lingering gaps. By creating structured lexical databases, Wordnets have served as a foundational support for a number of NLP tasks, yet they are constrained by sparse lexical relations and often lack domain specific vocabulary. Models with word context were contextualized thanks to their capacity to handle word context in a dynamic way, leading to a breakthrough in NLP tasks. But they depend heavily on tremendous amounts of data, and languages like these are typically under resourced. However, state of the art results have been achieved in many tasks using PLMs like GPT and its multilingual variants (e.g. mBERT, XLM R). However, the performance of such models is still biased as they are trained on datasets imbalance, i.e. they are trained on datasets where the languages with higher resources greatly outnumbers under-resourced language. Although powerful and capable of executing a variety of tasks, LLMs still struggle with resolving the intricacies of under-resourced languages comprising peculiar syntax as well as morphology due to limited computational resources in the quite resourced settings [20].

Even in the face of these challenges, several promising directions for making NLP more usable for under resourced languages remain. Transfer learning is rapidly becoming a major solution to help models leverage knowledge from high resource languages and generalize to low resource ones. Multilingual pretraining reduces bias by improving the representation of under-resourced languages and could be achieved by first doing pretraining on more diverse linguistic data during the training phase. New neural architectures for the morphological and non-standard syntax of these languages can also help to tackle their structural challenges. Less labeled data also presents the opportunity for the advancement of NLP for under-resourced languages through the use of unsupervised and self supervised learning [50]. Data scarcity problems can be greatly alleviated through collaborative effort like crowd-sourcing linguistic data and building open source resources. Theis will democratize access to NLP tools by creating lexicons, Wordnets and annotated corpora for under-resourced languages, producing high quality, diverse datasets involving native speakers, linguists, and data scientists. Lastly, strategies could be improved for fine-tuning on small datasets such as few shot learning and data augmentation, in order to bridge the performance gap and make use of small datasets [51].

To push NLP further towards inclusivity and fairness, challenges with these macro factors have to be addressed via creative strategies and community building, such that under resourced languages thrive along with the technological progress and linguistic diversity is retained.

# 5. Conclusion

We reviewed state of the art techniques in natural language processing while pointing out challenges in under-resourced languages. NLP boosted by Wordnets, contextualized models and other large pre-trained models has dramatically changed NLP tasks. Nevertheless, these advancements have been to the detriment of under-resourced languages which continue to suffer from data scarcity, linguistic diversity, and model pretraining bias. To address these challenges, we propose in this paper the use of unsupervised learning techniques, cross-lingual transfer learning, and open source datasets to under resourced languages. In addition, large scale language models can be combined into multilingual systems to achieve a more inclusive and representative models. These advances present a way to bridge the gap, so that under-resourced languages can take advantage of the ongoing pace of progress in the NLP arena.

As research in NLP progresses, it is critical that techniques are developed that emphasizes on linguistic diversity and inclusivity. Through a push for collaborative work between researchers, developers and policymakers, we believe there is possibility of building a future where all languages are fairly represented by technology and as a result, technology works for all communities and protects linguistic and language heritage.

# References

[1] X. Chen, H. Xie, X. Tao, "Vision, status, and research topics of Natural Language Processing," Natural Language Processing Journal, **1**, 100001, 2022, doi:10.1016/j.nlp.2022.100001.

[2] S. C. Fanni, M. Febi, G. Aghakhanyan, E. Neri, "Natural language processing," in Introduction to artificial intelligence, 87–99, Springer, 2023.

[3] B. Zhou, G. Yang, Z. Shi, S. Ma, "Natural language processing for smart healthcare," IEEE Reviews in Biomedical Engineering, **17**, 4–18, 2022, doi:10.1109/RBME.2022.3210270.

[4] Y. Zhou, X. Shen, Z. He, H. Weng, W. Chen, "Utilizing AI-Enhanced Multi-Omics Integration for Predictive Modeling of Disease Susceptibility in Functional Phenotypes," Journal of Theory and Practice of Engineering Science, **4**(02), 45–51, 2024, doi:10.53469/jtpes.2024.04(02).07.

[5] S. Huo, Y. Xiang, H. Yu, M. Zhu, Y. Gong, "Deep Learning Approaches for Improving Question Answering Systems in Hepatocellular Carcinoma Research," arXiv preprint arXiv:2402.16038.

[6] T. Shaik, X. Tao, Y. Li, C. Dann, J. McDonald, P. Redmond, L. Galligan, "A review of the trends and challenges in adopting natural language processing methods for education feedback analysis," Ieee Access, **10**, 56720–56739, 2022, doi:10.1109/ACCESS.2022.3177752.

[7] M. Abulaish, M. Fazil, M. J. Zaki, "Domain-specific keyword extraction using joint modeling of local and global contextual semantics," ACM Transactions on Knowledge Discovery from Data (TKDD), **16**(4), 1–30, 2022, doi:10.1145/3494560.

[8] J. Xiao, Y. Chen, Y. Ou, H. Yu, K. Shu, Y. Xiao, "Baichuan2-Sum: Instruction Finetune Baichuan2-7B Model for Dialogue Summarization," 2024 International Joint Conference on Neural Networks (IJCNN), 1–8, 2024, doi:10.1109/IJCNN60899.2024.10650513.

[9] D. M. Suresh, G. Vincent, C. Vijai, M. Rajendhiran, M. Com, A. Vidhyalakshmi, S. Natarajan, "Analyse Customer Behaviour and Sentiment Using Natural Language Processing (NLP) Techniques to Improve Customer Service and Personalize Banking Experiences," Educational Administration: Theory And Practice, **30**(5), 8802–8813, 2024.

[10] X. Wang, J. He, D. J. Curry, J. H. Ryoo, "Attribute embedding: Learning hierarchical representations of product attributes from consumer reviews," Journal of Marketing, **86**(6), 155–175, 2022, doi:10.1177/00222429211047822.

[11] Y. Gong, M. Zhu, S. Huo, Y. Xiang, H. Yu, "Enhancing Cybersecurity Resilience in Finance with Deep Learning for Advanced Threat Detection," arXiv preprint arXiv:2402.09820, 2024.

[12] L. Siddharth, L. Blessing, J. Luo, "Natural language processing in-and-for design research," Design Science, **8**, e21, 2022, doi:10.1017/dsj.2022.16.

[13] D. Cai, Y. Wang, L. Liu, S. Shi, "Recent advances in retrieval-augmented text generation," Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval, 3417–3419, 2022, doi:10.1145/3477495.3532682.

[14] S. Ranathunga, N. De Silva, "ome languages are more equal than others: Probing deeper into the linguistic disparity in the nlp world," arXiv preprint arXiv:2210.08523, 2022.

[15] D. Blasi, A. Anastasopoulos, G. Neubig, "Systematic Inequalities in Language Technology Performance across the World's Languages," arXiv preprint arXiv:2110.06733, 2021.

[16] J. Taylor, T. Kochem, "Access and empowerment in digital language learning, maintenance, and revival: a critical literature review," Diaspora, Indigenous, and Minority Education, **16**(4), 234–245, 2022, doi:10.1080/15595692.2020.1765769.

[17] R. Rocca, N. Tamagnone, S. Fekih, X. Contla, N. Rekabsaz, "Natural language processing for humanitarian action: Opportunities, challenges, and the path toward humanitarian NLP," Frontiers in big Data, **6**, 1082787, 2023, doi:10.3389/fdata.2023.1082787.

[18] D. Bzdok, A. Thieme, O. Levkovskyy, P. Wren, T. Ray, S. Reddy, "Data science opportunities of large language models for neuroscience and biomedicine," Neuron, **112**(5), 698–717, 2024, doi:10.1016/j.neuron.2024.01.016.

[19] M. Sigman, G. Cecchi, "Global organization of the lexicon," arXiv preprint cond-mat/0106509, 2001.

[20] B. Giordano, M. Prieur, N. Vuth, S. Verdy, K. Cousot, G. Serasset, G. Gadek, D. Schwab, C. Lopez, "Popcorn: Fictional and synthetic intelligence reports for named entity recognition and relation extraction tasks," Procedia Computer Science, **246**, 1170–1180, 2024, doi:10.1016/j.procs.2024.09.542.

[21] B. R. Chakravarthi, M. Arcan, J. P. McCrae, "Improving wordnets for under-resourced languages using machine translation," in Proceedings of the 9th global wordnet conference, 77–86, 2018.

[22] S. Vij, J. Juyal, A. Jain, D. Tayal, "Exploring WordNet® graphs for text summarization and sentiment analysis in Bengali speech," International Journal of Information Technology, 1–10, 2024, doi:10.1007/s41870-024-02285-z.

[23] C. Fellbaum, P. Vossen, "Challenges for a global wordnet," in Online Proceedings of the First International Workshop on Global Interoperability for Language Resources, 75–82, 2008.

[24] F. Stöhr, "Advancing language models through domain knowledge integration: a comprehensive approach to training, evaluation, and optimization of social scientific neural word embeddings," Journal of Computational Social Science, **7**(2), 1753–1793, 2024, doi:10.1007/s42001-024-00286-3.

[25] L. Cai, J. Li, H. Lv, W. Liu, H. Niu, Z. Wang, "Integrating domain knowledge for biomedical text analysis into deep learning: A survey," Journal of Biomedical Informatics, **143**, 104418, 2023, doi:10.1016/j.jbi.2023.104418.

[26] M. Zarzoura, From Abstract Syntax to Natural Language Addressing Natural Language Generation Challenges in Arabic Using GFWordnet as Lexical Resources., Ph.D. thesis, University of Gothenburg, 2024.

[27] A. Ramponi, "Language varieties of Italy: Technology challenges and opportunities," Transactions of the Association for Computational Linguistics, **12**, 19–38, 2024, doi:10.1162/tacl_a_00631.

[28] C. Zhao, M. Wu, X. Yang, W. Zhang, S. Zhang, S. Wang, D. Li, "A systematic review of cross-lingual sentiment analysis: Tasks, strategies, and prospects," ACM Computing Surveys, **56**(7), 1–37, 2024, doi:10.1145/3645106.

[29] A. Joshi, R. Dabre, D. Kanojia, Z. Li, H. Zhan, G. Haffari, D. Dippold, "Natural language processing for dialects of a language: A survey," ACM Computing Surveys, **57**(6), 1–37, 2025, doi:10.1145/3712060.

[30] M. A. Masethe, H. D. Masethe, S. O. Ojo, "Context-Aware Embedding Techniques for Addressing Meaning Conflation Deficiency in Morphologically Rich Languages Word Embedding: A Systematic Review and Meta Analysis," Computers, **13**(10), 271, 2024, doi:10.3390/computers13100271.

[31] S. Elayan, M. Sykora, "Digital intermediaries in pandemic times: social media and the role of bots in communicating emotions and stress about Coronavirus," Journal of Computational Social Science, 1–24, 2024, doi:10.1007/s42001-024-00314-2.

[32] V. D. Oliseenko, M. Eirich, A. L. Tulupyev, T. V. Tulupyeva, "BERT and ELMo in task of classifying social media users posts," in International Conference on Intelligent Information Technologies for Industry, 475–486, Springer, 2022, doi:10.1007/978-3-031-19620-1_45.

[33] A. Kantharuban, I. Vulić, A. Korhonen, "Quantifying the dialect gap and its correlates across languages," arXiv preprint arXiv:2310.15135, 2023, doi:10.48550/arXiv.2310.15135.

[34] A. Yusuf, A. Sarlan, K. U. Danyaro, A. S. B. Rahman, M. Abdullahi, "Sentiment analysis in low-resource settings: a comprehensive review of approaches, languages, and data sources," IEEE Access, 2024, doi:10.1109/ACCESS.2024.3398635.

[35] C. Wei, Y.-C. Wang, B. Wang, C.-C. J. Kuo, et al., "An overview of language models: Recent developments and outlook," APSIPA Transactions on Signal and Information Processing, **13**(2), 2024, doi:10.1561/116.00000010.

[36] B. Min, H. Ross, E. Sulem, A. P. B. Veyseh, T. H. Nguyen, O. Sainz, E. Agirre, I. Heintz, D. Roth, "Recent advances in natural language processing via large pre-trained language models: A survey," ACM Computing Surveys, **56**(2), 1–40, 2023, doi:10.1145/3605943.

[37] L. Wang, S. Chen, L. Jiang, S. Pan, R. Cai, S. Yang, F. Yang, "Parameter-efficient fine-tuning in large models: A survey of methodologies," arXiv preprint arXiv:2410.19878, 2024, doi:10.48550/arXiv.2410.19878.

[38] E. C. Zsiga, The sounds of language: An introduction to phonetics and phonology, John Wiley & Sons, 2024.

[39] A. Jiang, A. Zubiaga, "Cross-lingual offensive language detection: A systematic review of datasets, transfer approaches and challenges," arXiv preprint arXiv:2401.09244, 2024.

[40] Z. Chen, L. Xu, H. Zheng, L. Chen, A. Tolba, L. Zhao, K. Yu, H. Feng, "Evolution and Prospects of Foundation Models: From Large Language Models to Large Multimodal Models," Computers, Materials & Continua, **80**(2), 2024, doi:10.32604/cmc.2024.052618.

[41] J. Chen, Z. Liu, X. Huang, C. Wu, Q. Liu, G. Jiang, Y. Pu, Y. Lei, X. Chen, X. Wang, et al., "When large language models meet personalization: Perspectives of challenges and opportunities," World Wide Web, **27**(4), 42, 2024, doi:10.1007/s11280-024-01276-1.

[42] M. Z. Boito, V. Iyer, N. Lagos, L. Besacier, I. Calapodescu, "mhubert-147: A compact multilingual hubert model," arXiv preprint arXiv:2406.06371, 2024, doi:10.48550/arXiv.2406.06371.

[43] Y. Chang, X. Wang, J. Wang, Y. Wu, L. Yang, K. Zhu, H. Chen, X. Yi, C. Wang, Y. Wang, et al., "A survey on evaluation of large language models," ACM transactions on intelligent systems and technology, **15**(3), 1–45, 2024, doi:10.1145/3641289.

[44] T. H. Kung, M. Cheatham, A. Medenilla, C. Sillos, L. De Leon, C. Elepaño, M. Madriaga, R. Aggabao, G. Diaz-Candido, J. Maningo, et al., "Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models," PLoS digital health, **2**(2), e0000198, 2023, doi:10.1371/journal.pdig.0000198.

[45] S. Lankford, A. Way, "Leveraging LLMs for MT in Crisis Scenarios: a blueprint for low-resource languages," arXiv preprint arXiv:2410.23890, 2024, doi:10.48550/arXiv.2410.23890.

[46] A. Chaudhary, Automatic Extraction and Application of Language Descriptions for Under-Resourced Languages, Ph.D. thesis, Carnegie Mellon University, 2022.

[47] S. Sanjana, S. Kuranagatti, J. G. Devisetti, R. Sharma, A. Arya, "Intersection of Machine Learning, Deep Learning and Transformers to Combat Fake News in Kannada Language," in 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), volume 6, 2264–2270, IEEE, 2023, doi:10.1109/IC3I59117.2023.10398034.

[48] S. Liu, C. Chen, X. Qu, K. Tang, Y.-S. Ong, "Large language models as evolutionary optimizers," in 2024 IEEE Congress on Evolutionary Computation (CEC), 1–8, IEEE, 2024, doi:10.1109/CEC60901.2024.10611913.

[49] B. Haddow, R. Bawden, A. V. M. Barone, J. Helcl, A. Birch, "Survey of low-resource machine translation," Computational Linguistics, **48**(3), 673–732, 2022, doi:10.1162/coli_a_00446.

[50] K. Goswami, "Unsupervised deep representation learning for low-resourced," LASER, **77**, 79–7, 2012.

[51] T. ValizadehAslani, Y. Shi, J. Wang, P. Ren, Y. Zhang, M. Hu, L. Zhao, H. Liang, "Two-stage fine-tuning with ChatGPT data augmentation for learning class-imbalanced data," Neurocomputing, **592**, 127801, 2024, doi:10.1016/j.neucom.2024.127801.

ASTES

# A Study of the Digital Health Management Needs of the Elderly

Ya Gao[1*], Fatma Layas[2], Xiangyu Dong[1], Yijing Li[1], Jiayi Li[3]

[1] University of Wales Trinity Saint David, Wales Institute of Science of Art, Swansea, SA18PH, United Kingdom

[2] University of Wales Trinity Saint David, Assistive Technologies Innovation Centre, Wales Institute of Science of Art Swansea, SA18PH, United Kingdom

[5] Yanching Institute of Technology, Art Institute, Langfang, 065201, China

A R T I C L E   I N F O

A B S T R A C T

*The purpose of this paper is to explore the feasibility and development trend of utilizing smart medical technology for chronic disease health management in older people in the context of ageing at home. As the ageing society intensifies, the elderly population faces multiple health challenges, especially the management of chronic diseases. This paper analyzes the potential of smart medical technologies, such as remote monitoring, artificial intelligence, and the Internet of Things (IoT), to improve the efficiency and quality of health management for older people. By leveraging Maslow's Hierarchy of Needs Theory and Fogg's Behavioral Model, the article explores how to design smart health management products that meet the different health needs of older adults. In addition, the article discusses the barriers that the elderly population may encounter in accepting and using technology, such as the digital divide and technology adaptation issues, and proposes relevant coping strategies. Ultimately, the article concludes that with the continuous development of technology, smart healthcare technology will play an increasingly important role in geriatric health management, helping to improve the health status of older people, enhance their quality of life, and promote the innovation and development of social health management. The research in this paper provides new ideas for designing health management products for older people and supports the design and optimization of intelligent health management services.*

## 1. Introduction

China's ageing level is at the upper-middle level in the world, showing the characteristics of large population size and rapid ageing [1]. Subsequently, the contradiction between the explosive growth in demand for health services for the elderly and the continuous weakening of family care capacity has become increasingly prominent, and the development of digital health management is expected to be an effective way to solve this problem [2].

In the context of the global economic crisis in 2008, IBM proposed the concept of a "Smart Earth" for the first time, arguing

that the Internet of Things, the Internet, and intelligence together constitute the "Smart Earth" three elements[3].

In the same year, IBM released its "Smart Healthcare" solution in China, and since then ", Smart Healthcare" has been initially developed in China. With the rapid development of information technology, China's "smart healthcare" practice continues to deepen the development of digital health management, which is an important direction of the broader "smart healthcare" [4]. Intelligent health management uses a new generation of information, communication, artificial intelligence, bioinformatics, and other technological means to sense, analyze, and integrate the information from the three key links of health detection, health assessment, and health intervention so as to respond intelligently to the health needs of an individual or a group [5]. Although digital

health management offers solutions for older adults to cope with physical functioning, chronic diseases, and reduced socialization, older adults' difficulties in using digital health technology cannot be ignored. Personal, social, and technological factors impact the acceptance of digital health management among older adults. Therefore, it is of great significance to study the digital health management model for older people, which adapts to the trend of population ageing and meets the needs of older people. Combined with the concepts related to digital health management, and given the physiological and psychological characteristics of older people, the digital health management model for older adults can be briefly summarized as follows: under the premise of digital health technology adapted to ageing, health monitoring, assessment, and health interventions for older people can be implemented to efficiently, conveniently, and accurately satisfy the health needs of older people at all levels of older people's physiological, psychological, and social needs, and to improve older people's health level. For example, in the future, digital technologies such as artificial intelligence will be deeply integrated with the geriatric health service system, and older people will be able to obtain relevant health management services, such as health monitoring and assessment or guidance on medical treatment, by conversing with the system. The digital health management model will help older people integrate into digital city life and prompt them to improve their health. Digital health management technology also makes it possible for older people to stop being bothered by complex operating systems and to obtain personalized quality health services conveniently.

## 2. Technologies and Applications Related to Digital Health Management for Older People

Current digital health management technologies for older people include remote monitoring, mobile health, the Internet of Things, smart homes, and artificial intelligence.

Telemonitoring systems help patients collect health data online at home and transmit it to health centers. The telemedicine model provides critical monitoring support for chronic diseases and older people. Wang et al. developed a community-based health monitoring system for older people. The study first used relevant equipment to obtain and record relevant health information such as daily activities, continuous vital signs, and gait of older people, after which the decision support system utilized advanced data mining techniques to count the significant changes in the data and accordingly sent alerts to older people and their families and their The decision support system utilizes advanced data mining techniques to count significant changes and accordingly alerts older people and their families and caregivers to take appropriate interventions to prevent deterioration of health conditions[6]. Remote monitoring technology provides significant support for healthy, safe, and independent living, especially for older adults with chronic conditions such as cardiopulmonary disease, asthma, and heart failure, and has great promise for the future.

The World Health Organization defines "mobile health "as a service delivery method that uses cell phones, tablets, and wearable devices to provide medical support. Through the mHealth platform, health information can be continuously delivered from the patient's end to the doctor, and the doctor's solutions can be delivered to the patient so that problems arising in

the patient's body can be judged and solved in advance [7]. Mobile health apps play a significant role in the physical assessment of older people. Silva et al. have developed an app called "Geriatric Assistant" as a practical guide for healthcare professionals to assess older people's health and access up-to-date information comprehensively [8]. Wearable devices have also developed rapidly in recent years. Yu et al. developed a deep-learning model based on wearable device data to monitor falls among older people [9]. Zhong et al. investigated gait assessment applications for older adults to provide favorable support for analyzing regularity, symmetry, and variability of gait length in older adults [10].

The Internet of Things (IoT) connects things to things and things to people through various devices [11]. With the continuous reform and innovation of device functions, IoT technology is deeply developed in the health field to assist in meeting the health management needs of older people. Liu et al. designed a health promotion system for older people using IoT technology, which can organize the long-term dietary and exercise records of older people and assist older people in completing their personal nutritional assessments and health management. IoT is widely used in smart cities, public services, smart homes, life health, and personal care[12]. In addition, the development of IoT technologies such as blockchain technology, tactile internet, and nano-internet is anticipated.

Smart home technology provides a degree of digital linkage or living experience, creating a unique home for the user with sensors and actuators configured together. With the continuous development of smart home technology, smart home products in health and ageing are also emerging. However, while increasing the experience, there is no compelling evidence that smart homes significantly affect the treatment of diseases and the prevention of incapacitation [13].

Artificial intelligence, as a technology, is mainly applied to medical robots to make intelligent machines react similarly to humans through the logical judgment of autonomous intelligence. Current intelligent machines include features such as face recognition, intelligent speech, and deep learning. In addition, artificial intelligence plays a vital role in serving older people. Artificial intelligence software helps older persons sift through exercise, diet, and other health information. To a certain extent, robotic pets can reduce the sense of loneliness among older persons, and chatbots can communicate with older persons and remind them of the time to take medication and have regular medical check-ups [14]. In the future, AI will play a more prominent role in senior living needs. With the continuous progress of science and technology, artificial intelligence technology is gradually reaching its potential great value in the medical field, artificial intelligence technology is gradually realizing its potential great value in the medical field. Artificial intelligence is an important driving force for the new round of technological revolution and industrial change. It is an important driving force for the new round of scientific and technological revolution and industrial change. It is a new technical science that researches and develops theories, methods, technologies and application systems for simulating, extending and expanding human intelligence. It is a new technological science that researches and develops theories, methods, technologies and application systems for simulating, extending and expanding human intelligence. Compared with

traditional management strategies, the technical support provided by AI technology offers a more efficient and convenient way to monitor and treat chronic diseases. The characteristics of AI algorithms enable them to process massive amounts of data to more accurately detect hidden patterns or trends, which is an extremely critical feature for predicting disease progression or evaluating the effectiveness of treatment [15]. Compared to manual labour that requires rest, AI systems can operate 24/7 without interruption, continuously providing real-time feedback and advice to doctors and patients, further improving the efficiency of management. AI-based prediction models also help doctors develop more personalized treatment plans to better meet the individual needs of patients.

## 3.   Impact of Digital Health Technologies on Older Adults

The design has been transformed into a technological design when design activities based on computer network technology and virtual reality technology enter people's lives. Technology has been controversial since the beginning of the industrial age. Optimists believe that technology is the greatest invention of the age, creating things that did not exist in the world and greatly enriching people's lives. On the contrary, pessimistic people believe that technology goes against ethics and morality and even leads to the decline of culture. Of course, in today's age of information technology, people's perception of technology is not simply black and white, especially when it comes to intelligent medicine.

Positive Impacts: Due to the deep plumbing of information technology, we cannot ignore its many positive impacts. On the one hand, IT provides strong technical support for safe and healthy healthcare. On the other hand, it also provides more diverse disease prevention and treatment services for older people.

It is an inevitable trend for digital health technology to provide intelligent services for older people, and it is also a positive role of technology in healthcare services. Telemedicine services enable older people to communicate with their doctors online, reducing to a certain extent the cases in which older people cannot seek medical treatment even for various reasons. This is not only helpful to older persons living in remote areas but also provides vital support to those in poor health and unable to seek medical care in a timely manner.

There is evidence that telemedicine positively impacts the health management of homebound older adults with diabetes, including some reduction in cognitive decline, mortality, hospitalization, and healthcare costs, and may increase disease-related knowledge, adherence, and self-efficacy. On the other hand, healthcare professionals conduct two-way video disease monitoring and health management guidance with home-bound elderly patients through the Internet platform. They can also carry out popularization education, dietary monitoring, rehabilitation guidance, and health risk and medication adherence assessment[16]. Most elderly home-based caregivers must assist older people in their daily lives, monitor their signs, or address emergencies. However, most caregivers do not receive formal training in caregiving. As a result, home-based caregivers need more medical information training and support tools to facilitate stress management and improve their coping skills. Mercy is the world's first virtual medical center, which uses advanced

technology to provide telemedicine services, including remote round-the-clock health consultations, emergency care, and home monitoring. Healthcare professionals can collect detailed data from seniors through video monitors to capture some of their sudden symptoms so that treatment can be taken [17]. Telemedicine solves, to some extent, the geographical problem for both healthcare providers and home-based caregivers, as healthcare providers can not only provide psychosocial interventions, training and support to older caregivers (including family members and other informal caregivers) through telemedicine platforms but also offer comprehensive geriatric care programs to caregivers [18].

**Negative impact:** The use of technology has not been smooth sailing for older people. The conflicts and contradictions that have erupted among older people in the Internet era have forced society to ponder whether the times are moving so fast that we are ignoring a part of the population that technology has forgotten. Does the rapid update of technology have no negative impact? The answer is no.

The dilemma older people face in the Internet era is the "digital divide" phenomenon in the academic world. Scholars generally believe that the fundamental dilemma that prevents older people from using intelligent technologies in a learning-oriented manner is not their lack of interest in or rejection of new technologies but rather their state of existence, conditions, and environment, which leads to their objective "vulnerability" in the face of the promotion and use of intelligent technologies[19]-[20].

The first is vulnerability at the physiological level. This condition is an objective problem for older people when exposed to innovative technologies. With the continuous development of medical technology, there is no doubt that human life expectancy has gradually increased. However, as older people age, the decline in cognitive functions such as vision, hearing, and touch, as well as in cognitive and thinking skills such as attention and memory, remains irreversible. Moreover, these problems become more pronounced with age. This is even though relevant research suggests that most older people are, to some extent, internally or externally motivated to learn intelligent technologies [21]. Differences in educational backgrounds and work experiences make older people, to a certain extent, reflect specific differences. However, the primary trend is similar, i.e., "slow walkers" in the "fast" era of intelligence. Media theory suggests that digital and intelligent technologies are essentially seen as extensions of people's perceptions and have a symbiotic relationship with human beings, being "part of humanity [22]". However, the increasing age of older people and the constant iteration of technology have made the predatory disadvantage of older people even more exceptionally pronounced.

The second is vulnerability at the cultural level. The experience and exposure that older people have accumulated over the long years is a priceless asset, yet it is undeniably a form of entrenchment and constraint for them. This leads them to favor traditional and familiar technologies. They feel a vague sense of alienation in the fast-developing smart era, which leads to a sense of "fear" or "rejection" when facing intelligent technologies. This situation may lead to a lack of interest in and exposure to innovative technologies [23]. Therefore, compared with young people, it is difficult for older people to develop intelligent

technology information literacy and overall cultural atmosphere [24].

In addition, vulnerability has a social dimension. Social support from family members plays a crucial role in helping older people cross the digital divide [25]. However, the role of family support in the intelligent technology enhancement of older people has not been fully utilized; on the one hand, due to the scarcity of time for children and other family members, when older people have digital needs and ask for help from family members, the family members may not have enough patience to teach older people in-depth. Older people have not grasped the essentials of learning digital technology, which, to a certain extent, may deepen their rejection of digital products. On the other hand, it is due to the "generation gap" that limits the willingness and initiative of older people to seek help [26]. The gradual withdrawal of older people from various social relationships has led to a gradual scarcity of their social roles, such as peer and social relationships, and a weakening of their social status, making the conditions for older people to take the initiative in obtaining help weaker and weaker, which will make older people not have enough ability to obtain information related to digital technology, and the tendency to "weaken themselves" is constantly emerging in the use of intelligent technology [27]-[29].

The topic worth exploring is how digital technology can be adapted to an ageing society and practically help make life more convenient for older people. Therefore, while bridging the digital divide, we should accept that ageing is an objective reality of social development. It is undeniable that the development of technology has provided diverse services for older people. However, at the same time, technological development cannot relentlessly leave older people behind. Therefore, we need to start from the perspective of the needs of older people, really see their expectations, and appropriately meet their needs so that intelligent technology can serve the lives of older people more reasonably.

## 4. Needs analysis of older people

### 4.1. Hierarchy of needs of older people

Given the diverse and complex needs of elderly patients with chronic diseases, Maslow's needs theory will provide us with directions. Abraham Maslow, an American psychologist, proposed Maslow's theory of needs in 1943, dividing human needs into five levels from low to high: physiological needs, safety needs, social needs, respect needs and self-actualization needs (as shown in the figure below). Everyone has needs, and when low-level needs are satisfied, people pursue higher-level needs. When multiple needs are unsatisfied, people will first pursue the most urgent needs [30].
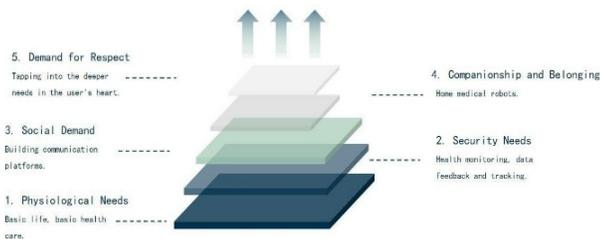


Figure 1: Maslow Hierarchy of Needs

Physiological needs are the most basic necessary for survival. If physiological needs are not met, the lives of older persons are threatened. Therefore, technology supports the basic life and basic medical care of the elderly. Digital health technology uses products or services to solve the fundamental problems of life and medical care for the elderly.

Security needs include personal safety, health protection, moral security, family security and property ownership. Among them, health protection is the foundation of other security needs of older people. In other words, only with a healthy body can older people start other activities. Older adults need health monitoring due to their declining physical functions. Therefore, physiological signs such as heart rate, blood glucose, blood pressure, sleep, etc., can be monitored with data feedback and tracking by using sensing technology to report the physical condition through several critical indicators so that timely treatment and resuscitation can be carried out in dangerous moments to improve the efficiency of medical treatment.

Social needs are higher-level needs, including love, friendship and affection. Compared with physical needs, emotional needs are more detailed. Everyone is in a specific social environment and wants to be cared for. Specific social needs are unavoidable in order to avoid self-enclosure. However, their social needs differ from those of young people; they mostly socialize by organizing activities online and gathering offline, and their channels to broaden their social circle are concentrated offline, and network socialization is just a tool for contacting their feelings. Therefore, the role of community peers in the self-health management of older people is self-evident. Innovative healthcare platforms can connect patients with groups with similar health conditions, promote mutual support and group communication, and increase the social participation of older people. Nowadays, many products have APP services, in which a communication platform can be built for older people, bringing together older people with the same disease who can discuss their daily lives and provide each other with experiences and emotional exchanges.

Respect the need for self-expression of older people. Everyone is unique and has the right to express themselves; of course, older people are no exception. Digitalization has given rise to the development of short videos, and the cost of self-expression has been dramatically reduced, empowering the general public, including older people, with more voice. Used for self-expression is an affirmation of oneself. Digital platforms have brought about a noticeable shift in the daily lives of older adults, providing a platform for those brave enough to express themselves.

Older people build confidence in self-health management, which is conducive to leading a decent life. Therefore, we can make full use of artificial intelligence and sensing technology to enhance the "proactivity" of products, meet the functional needs of users, and even internalize them as part of the bodily functions of the elderly, forming a natural and hidden interaction. In this way, the psychological needs of older people who do not want their physical ageing and defects to show, do not want to be looked at differently by others, and desire to live a decent life like ordinary people can be satisfied.

## 4.2. Linking Smart Technology to the Needs of Older Adults

In addition to meeting the needs of older people, it is also essential to strengthen the relationship between older people and technology, which helps to enhance their willingness to use innovative products. Therefore, behavioral design is also needed. Through behavioral design, we can stimulate the interest of older people in using innovative products and guide older people who are already using intelligent products to use them more intensively.

In 1930, Harvard psychologist B.F. Skinner created a Skinner box experiment to study how rats respond to rewards. In a box with a control lever, the rats were given food to drop whenever they pushed the lever, a reward that led to the rats quickly learning the skill of pushing the lever. Skinner speculated that if the rewards were designed well, human behavior could be guided by what is now known as Behavioral Design. Professor B.J. Fogg of Stanford University applied behavioral design with computer software and the Internet to provide theoretical support for studying user behavior [31]. Fogg believed that behavior is influenced by motivation, capability cost, and triggers, i.e., the Fogg Behavioral Model (shown below) was produced. The Fogg Behavioral Model shows that the occurrence of behavior needs to satisfy the three critical points of motivation, capacity cost, and triggers, which are favorable conditions for the occurrence of a behavior.
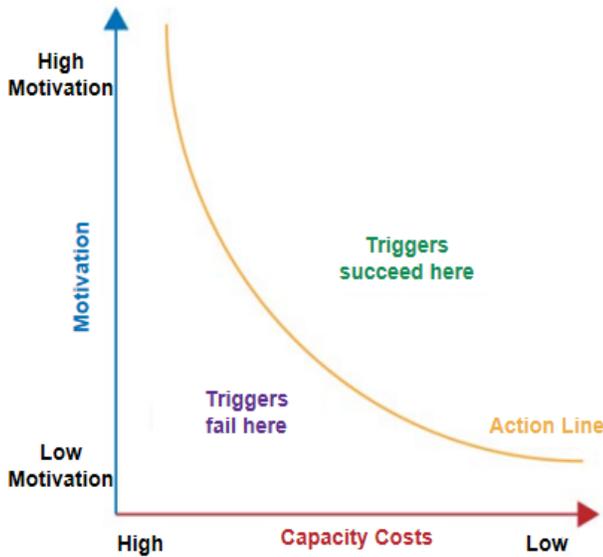


Figure 2: Fogg Behavior Model

① Increase the motivation to use the product: Motivation can be divided into direct and indirect. The user's willingness to use this product is low. Suppose we want to fully integrate ageing and digitalization and make digitalization fully integrated into the lives of older people. In that case, we need to start with the people or environment around older people and indirectly influence the willingness of the elderly group to use intelligent products. Recommendations from friends and relatives are an excellent way to gain motivation. Due to the obvious social needs of older people, getting likes and comments from their friends and relatives on technology products will increase their willingness to use them and achieve the purpose of communication and socialization. In addition, the "family" function also makes online socializing more

cohesive. Among them, the "regional family" has very regional attributes and can be accurate to the provinces, cities and streets where the elderly users can not only expand offline social networking but also increase the sense of belonging of older people in use.

② Reduce the cost of using digital products: Motivation to meet the factors and the ability to cost are also indispensable.

a. Reduce the user's time cost, i.e., improve the product's responsiveness and give positive feedback during the waiting process to increase the product's fault tolerance. Otherwise, older people may perceive that they have made a mistake and thus be demotivated to use the product.

b. Reduce learning costs. Intelligent products with clear functional logic and simple interaction may attract many users. For example, in Dou Yin, videos can load automatically, and users can see updated videos by simply sliding up or down. Older people can access their favorite content by simply moving their fingers, and the learning cost is meagre, attracting many elderly users.

③ Increase the triggers to use the product: although the user has generated the motivation to use the product, the capacity cost is not high, but if they need a specific behavior to occur smoothly, the catalyst between the elderly user and the product is also significant, that is the triggers (as shown in the figure below).
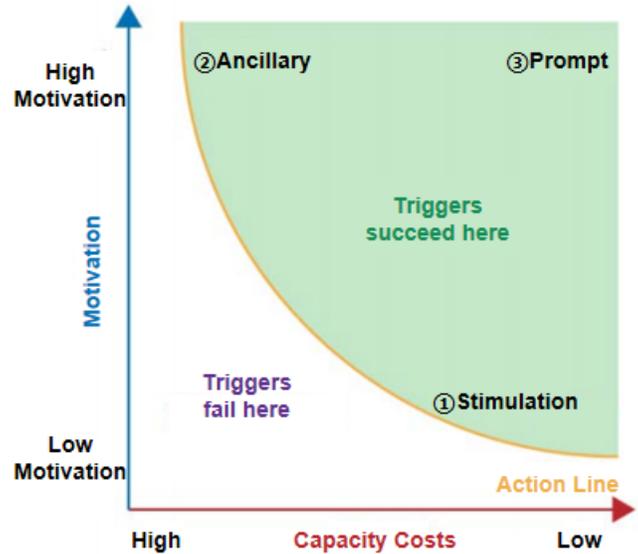


Figure 3: Triggering Factor

a. Reward stimulus is the most common trigger. In the case of a product with low-capacity cost but weak user motivation to use it, reward stimulation affects the user's usage behavior. Getting unexpected rewards when the user uses the product can increase the user's sense of surprise about the product and improve the stickiness of using the product. For example, in the case of Pinduoduo, the help of friends can enable users to obtain a sum of money. Users can get rewards through meagre capacity costs, so users who are not too willing to use the product also develop the behavior of using the product.

b. Assistance and product fault tolerance. When older users are motivated to use the product, but the ability cost is high, appropriate instructions and guidance can reduce the learning cost to a certain extent. "Kaixin Xiaoxiao Le is one of the few popular games among the elderly. In addition to the friendly image design, more importantly, every new element or character in the game will guide the user on how to play, and when a new activity appears, it will also guide the user on how to get the prize. The prompt signals in the product can also communicate with the user to avoid the situation where the user suddenly encounters a problem that leads to an inability to operate.

## 5. Conclusion

To focus on technology is to focus on the future. Under the premise of comprehensively judging the situation of population ageing and chronic diseases, an in-depth analysis of the positive and negative impacts of intelligent medical technologies will lay the foundation for finding solutions to innovative medical technologies to better meet the needs of older adults in the future. Elderly health management services and products relying on remote monitoring, artificial intelligence and other information technologies mean there are new ways and contents for elderly health management services. In the self-health management of older adults, we need to pay attention to the different levels of health needs and support them.

Guided by the goal of comprehensively improving the health of the elderly population, Maslow's hierarchy of needs theory provides theoretical support for the study of self-health management of older adults. Fogg's behavioral model provides a reference for the design of intelligent health products. This study uniquely combines Maslow's and Fogg's theoretical frameworks to explore and address the particular challenges faced by older adults in chronic disease management. By combining insights from behavioral psychology with cutting-edge technology applications, this study provides innovative perspectives for designing inclusive, personalized health management solutions. It is believed that with the continuous acceleration of the process of scientific and technological innovation, the application of science and technology in the field of digital health will become more and more extensive and will continue to meet the multi-level and diversified needs of the elderly chronic disease group, thus improving the efficiency of health management and quality of life of older people [32].

## References

[1] Z.P. Ren, "China ageing report," Development Research, 40(2), 22–30, 2023.

[2] E. Vidal, "Digital literacy program: Reducing the digital gap of the elderly: Experiences and lessons learned," in 2019 International Conference on Inclusive Technologies and Education, IEEE, San Jose del Cabo, Mexico: 117–1173, 2019, doi:10.1109/CONTIE49246.2019.00030.

[3] IBM, Smarter planet, 2008.

[4] M. Chen, "New progress and prospect of smart healthcare in China in 2021," China Medical News, 36, 6–7, 2021.

[5] Q. Zeng, X.Y. Gao, S.Z. Bai, "Theory and practice of smart health management," Chinese Journal of Health Management, 16, 3–6, 2022.[6] H. Wang, Y. Zhao, L. Yu, et al., "A personalized health monitoring system

for community-dwelling elderly people in Hong Kong: design, implementation, and evaluation study," *J. Med. Internet Res.,* vol. 22, p. e19223, 2020.

[6] H. Wang, Y. Zhao, L. Yu, J. Liu, I.M. Zwetsloot, J. Cabrera, K. L. Tsui, "A personalized health monitoring system for community-dwelling elderly people in Hong Kong: Design, implementation, and evaluation study," Journal of Medical Internet Research, 22(9), e19223, 2020, doi:10.2196/19223.

[7] R.P. Searcy, J. Summapund, D. Estrin, J.P. Pollak, A. Schoenthaler, A.B. Troxel, J.A. Dodson, "Mobile health technologies for older adults with cardiovascular disease: Current evidence and future directions," Current Geriatrics Reports, 8(1), 31–42, 2019, doi:10.1007/s13670-019-0270-8.

[8] S. Silva, R. Felgueiras, I.C. Oliveira, "Geriatric helper: An mHealth application to support comprehensive geriatric assessment," Sensors, 18(4), 1285, 2018, doi:10.3390/s18041285.

[9] S. Yu, Y. Chai, H. Chen, R.A. Brown, S.J. Sherman, J.F. Nunamaker, "Fall detection with wearable sensors: A hierarchical attention-based convolutional neural network approach," Journal of Management Information Systems, 38(4), 1095–1121, 2021, doi:10.1080/07421222.2021.1990617.

[10] R. Zhong, P. L.P. Rau, "A mobile phone–based gait assessment app for the elderly: Development and evaluation," JMIR mHealth and uHealth, 8(5), e14453, 2020, doi:10.2196/14453.

[11] A. Panarello, N. Tapas, G. Merlino, F. Longo, A. Puliafito, "Blockchain and IoT integration: A systematic survey," Sensors, 18(8), 2575, 2018, doi:10.3390/s18082575.

[12] M. Shakeri, A. Sadeghi-Niaraki, S. M. Choi, S.M.R. Islam, "Performance analysis of IoT-based health and environment WSN deployment," Sensors, 20(20), 5923, 2020, doi:10.3390/s20205923.

[13] L. Liu, E. Stroulia, I. Nikolaidis, A. Miguel-Cruz, A. Rios Rincon, "Smart homes and home health monitoring technologies for older adults: A systematic review," International Journal of Medical Informatics, 91, 44–59, 2016, doi:10.1016/j.ijmedinf.2016.04.007.

[14] J. Rong, X. Ji, X. Fang, M. H. Jee, "Research on material design of medical products for elderly families based on artificial intelligence," Applied Bionics and Biomechanics, 2022, 1–6, 2022, doi:10.1155/2022/7058477.

[15] A.M. Fischer, A. Varga-Szemes, M. Van Assen, L.P. Griffith, P. Sahbaee, J.I. Sperl, J.W. Nance, U.J. Schoepf, "Comparison of artificial intelligence–based fully automatic chest CT emphysema quantification to pulmonary function testing," American Journal of Roentgenology, 214(5), 1065–1071, 2020, doi:10.2214/AJR.19.21572.

[16] C.L. Walker, M. Kopp, R.M. Binford, C.J. Bowers, "Home telehealth interventions for older adults with diabetes," Home Healthcare Now, 35(4), 202–210, 2017, doi:10.1097/NHH.0000000000000522.

[17] L. Klingensmith, L. Knodel, "Mercy virtual nursing: An innovative care delivery model," Nurse Leader, 14(4), 275–279, 2016, doi:10.1016/j.mnl.2016.05.011.

[18] N. C. Chi, G. Demiris, "The roles of telehealth tools in supporting family caregivers: Current evidence, opportunities, and limitations," Journal of Gerontological Nursing, 43(2), 3–5, 2017, doi:10.3928/00989134-20170111-04.

[19] J. Wu, "Study on the current situation and influencing factors of China's elderly people's internet use: Analysis based on CGSS 2017 data," Scientific Research on Aging, 9(9), 43–58, 2021.

[20] S.J. Song, Research on digital divide problems and countermeasures for the elderly in Dalian city, Liaoning Normal University, 2021, doi:10.27212/d.cnki.glnsu.2020.001653.

[21] W.C. Su, Z.P. Lu, Z.X. Wang, "Preserving the dignity of elders' choices:

A conceptual model of autonomous behavior in the use of digital technologies for older adults,＂ Library Forum, (8), 86–95, 2021.

[22] Y.L. Wang, L. Z., ＂Analysing the influencing factors and countermeasures of the digital divide among the elderly in the digital age,＂ Journalism Probe, 9, 126–128, 2021.

[23] F. Lin, A study of the digital generation gap in Chinese families in the context of ageing, Shenzhen University, 2018.

[24] X. Xu, The impact of the internet on the continued socialization of older people, Dongbei University of Finance and Economics, 2013.

[25] W.P. Zhang, J.H. Fan, ＂The construction of a social support system for the digital divide in old age,＂ Scientific Research on Aging, 7(2), 63–70, 2019.

[26] Z.C. Du, ＂Revisiting the adherence to a positive view of gerontological education — and discussing the social de-personalization of the individual, life education,＂ The Seniors University, 5, 22–26, 2009.

[27] C.X. Huang, ＂Status, challenges and countermeasures of the digital divide in old age,＂ People＇s Forum, (29), 126–128, 2020.

[28] B. Lin, ＂Difficulties and priorities in the governance of digital poverty among the elderly,＂ People＇s Forum, (29), 129–131, 2020.

[29] B. Yang, D.C. Jin, ＂The digital divide in the elderly: Manifestations, motivations, and paths to bridging,＂ Central States Journal, (12), 74–80, 2021.

[30] A.H. Maslow, Maslow＇s humanistic philosophy, Jilin Publishing Group Co., Ltd.: 35–38, 160–168, 2013.

[31] B. Fogg, ＂A behavior model for persuasive design,＂ in Proceedings of the 4th International Conference on Persuasive Technology, ACM, Claremont California USA: 1–7, 2009, doi:10.1145/1541948.1541999.

[32] H.L. Zhu, ＂Integrating technology in China to help the elderly: A paradigm shift,＂ Communication of IIMA, 3, 95–106, 2015.