# Navigating the new frontier: large language models and their implications for education

**Laila Boullous[1,2], Mustapha Hain[1], Adil Chergui[1], Brahim Elbhiri[2]**
[1]Laboratory of Artificial Intelligence and Complex Systems Engineering (AICSE), National School of Arts and Crafts (ENSAM), Hassan II University, Casablanca, Morocco
[2]Laboratory Smartilab, Moroccan School of Engineering Sciences, Rabat, Morocco

**Article Info**

***Article history:***

Received Nov 3, 2024
Revised Mar 7, 2026
Accepted Apr 22, 2026

***Keywords:***

Applications
Artificial intelligence
Challenges
Comparative analysis
E-learning
Generative models
Traditional language models

**ABSTRACT**

This survey characterizes the contributions of large language models (LLMs) to technology enhanced learning by relating their capabilities to actual educational functions, making comparisons with traditional models of language. The contributions for this study are: i) introduce an education centered taxonomy that classifies LLM use by four key functions personalization and adaptivity, assessment and evaluation, profiling and prediction, and intelligent tutoring with illustrations from deployed systems and tools; ii) give a domain-based comparison of where LLMs outperform traditional models (sentiment analysis with sarcasm, context-aware question answering, and abstractive summarization) and why those advantages will mean something to e-learning practice; iii) synthesize six cross-cutting risks, including computational cost/carbon, privacy, bias and hallucination, labor displacement, interpretability, and the limits of human-like judgment, and provide practical design/research implications; and iv) report on a transparent review protocol that got the initial corpus down to 50 key articles, allowing for modifications and future updates from other interested researchers. In sum, the discussion about LLMs in education has been pushed past the broad strokes to a situation where there is a comprehensive vocabulary for what LLMs can do, and how they may or may not responsibly improve learning experiences, educator workflows, and systems/learning design in e-learning.



***Corresponding Author:***

Laila Boullous
Laboratory of Artificial Intelligence and Complex Systems Engineering (AICSE)
National School of Arts and Crafts (ENSAM), Hassan II University
Casablanca 20700, Morocco
Email: laila.boullous-etu@etu.univh2c.ma

## 1. INTRODUCTION

In recent years, there has been a remarkable revolution in artificial intelligence (AI), particularly in the field of large language models (LLMs) [1]. LLMs, which possess immense computational power and have been trained on enormous training datasets, have transformed generation and natural language processing (NLP) problems [2]–[4]. Therefore, it's important to understand LLMs what: they are, how they came to be, the various types, and their advantages over traditional language models.

The revolutionary upheaval emerged with the publication of the transformer architecture, which utilizes a self-attention mechanism to identify long-range dependencies [1], [5], [6]. Building on this, models like bidirectional encoder representations from transformers (BERT) and generative pre-trained transformer (GPT)-1 were developed to comprehend context and produce human-like text [7]–[9]. Specifically, models such as GPT-3 and GPT-4 have pushed the boundaries of NLP tasks using unsupervised pre-training and supervised fine-tuning [10]–[12]. Furthermore, specialized domain-specific models have emerged in fields like healthcare (medical pathways language model (Med-PaLM)) and finance (financial generative pre-trained transformer (FinGPT)) to address task-specific limitations. Therefore, it's important to understand LLMs: what they are, how they came to be, and their advantages over traditional language models [13], [14].

This paper investigates the introduction of generative models, analyzes LLMs in detail, and highlights their advantages over traditional models, which often struggle with complex items like metaphors and sarcasm. It also examines the limitations and challenges of LLMs, such as computational demands, data privacy, and "hallucinations". Furthermore, it explores the implications and ways LLMs can aid the educational context by facilitating personalization, adaptivity, and intelligent tutoring.

## 2. METHOD

The review process was undertaken with a systematic step-by-step approach to ensure that all research questions and other examined topics on LLMs and their application to education were comprehensively reviewed. The process, as outlined in Figure 1, began with a literature search phase. A broad range of academic sources was uncovered using specific keywords such as, "LLMs", "large language models", "generative AI", "education", and "e-learning". This search resulted in a considerable number of articles which were reviewed iteratively based on their relevance, academic rigor, and contributions to topic. After this review, an initial set of 50 articles were identified and selected that were deemed the core for the study. These articles established a base of knowledge on the technical workings of LLMs and the implications of LLMs beyond the technical aspects within education.

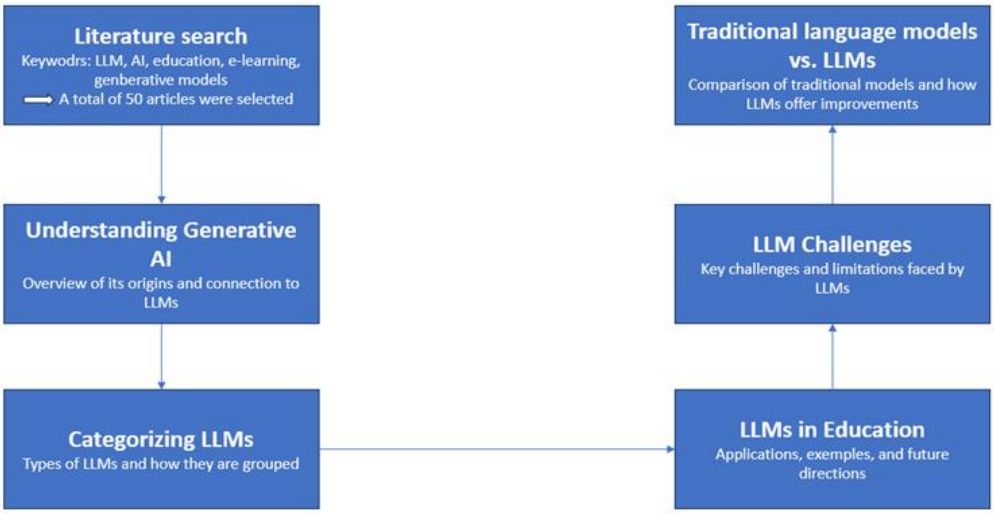


Figure 1. Overview of the research methodology for exploring LLMs and their educational applications

Following this, research entered a phase of general inquiry where research questions were developed to promote the detailed exploration of the topic. The inquiry began with investigating generative AI as an umbrella term to understand aspects such as its origins, when generative AI started to gain traction, and how its growth and development lead to the emergence of LLMs. This inquiry also examined a historical perspective of AI, in order to recognize key development periods which, contribute to the evolution. The questions being asked during this inquiry also assisted to clarify how LLMs fit into the overall variable landscape of AI technologies that can be considered for specific inquiry movements. The third step of the methodology presented a closer analysis of LLMs and their categorization. As there are a multitude of LLMs in today’s marketplace, the research aims to explore the categorization of LLMs, or in other words how LLMs are grouped. The process consisted of categorizing LLMs and analyzing how they are grouped by size, architecture, and application.

Once the categorization was complete, the research began comparing traditional language models to LLMs. The comparative methodology aims to expose the limitations of traditional language models, which have dominated the field, and how LLMs have addressed those limitations. In particular, comparison focused on LLMs' abilities to handle larger data sets and perform more complex tasks while yielding more accurate results. Utilizing a comparative approach elucidates the progress in the field of second language acquisition and highlights why LLMs necessitate merit as a new and innovative approach in the field of education.

After illustrating the strengths of LLMs, the comparison shifted to limitations and challenges of LLMs. While LLMs are clearly an enhancement to second language acquisition, LLMs should not be seen as perfect. This part of the review proposes a critical evaluation of the challenges of LLMs. Examining the challenges presents a balanced understanding of the potential and limitations of LLMs.

The last portion of the methodology looks more closely at LLMs and education. The review raises issues of how LLMs are perceived in educational environments and whether there is a focus on e-learning environments. Finally, this analysis points to possible outcomes for LLMs and also provides issues in LLMs where development and research are needed to better students and educators with LLM tools.

## 3. EVOLUTION OF GENERATIVE MODELS IN ARTIFICIAL INTELLIGENCE

The story of the generative model, which is illustrated in Figure 2, begins when Alan Turing proposed Turing machines in 1936 when he built the models capable of imitating and reproducing linear data. Hidden Markov models (HMM) began to arise in 1960 that provided a new probabilistic-based process to infer and predict sequences. 1982 marks the introduction of recurrent neural networks (RNNs), unlocking the ability to remember previous inputs. Afterwards, the long short-term memory (LSTM) was suggested in 1997, resulting in facilitating various tasks, including time series prediction, by overcoming the challenges related to long-term dependencies.

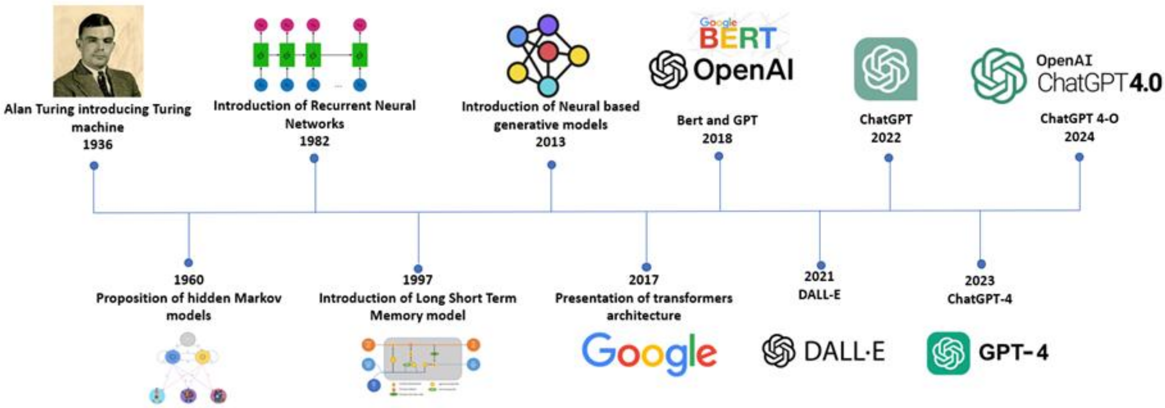


Figure 2. A timeline of generative AI models starting from Turing machines in 1936 to modern GPT architectures

The rise of transformers started with the proposition of a new framework in 2013 that helped generate new data based on a training set; the type of generated data varies from textual, imagery, audio, and more. Using neural-network-based generative models such as the variant auto-encoders (VAEs) and generative-adversarial networks (GANs). The revolutionary upheaval emerged with the publication of the paper “Attention is all you need” in 2017 [2], where Google presented the architecture of transformers and their power that lies in the self-attention mechanism and its ability to identify long-range dependencies and trace connections within the data [1].

The following year the suggested framework was implemented into LLMs. Google introduced BERT, an extensive language model, its power lies in its proficiency in comprehending the context of words in a sentence [7]. OpenAI presented its LLM GPT with its ability to produce human-like text [8].

With the apparition of DALL-E in 2021, the world knew that transformers capabilities extend NLP and can be used for image generation from textual prompts. OpenAI, founded in December 2015, has contributed mainly to the advancement of LLMs and influenced the landscape of AI research and applications. They focused on building generative models, starting from GPT-1 to launching ChatGPT in 2022. The popularity of LLMs has increased exponentially after the launch of ChatGPT, as it has gained millions of users over a few weeks after its release thanks to high performance with unseen tasks based on its ability to generalize from taking a task description as input and a few examples. It doesn’t require any task-specific training, which opens the door to a new era in AI, introducing what we refer to as artificial general intelligence [10].

GPT models combine unsupervised pre-training and supervised fine-tunning to generate human-like answers. GPT-3 is a version of GPT models built with 175 billion parameters, trained on a large corpus containing of textual data from web pages, books, articles, and social media [3], [4], [10]. OpenAI introduced GPT-4, an advanced version of GPT, in March 2023 with a massive number of parameters that goes up to 1.76 trillion parameters; it can process both image and text inputs. GPT-4 capabilities go beyond generating human-like text, this model family has further pushed the performance results in various NLP tasks [10].

The latest version of ChatGPT is ChatGPT-4o (the last ‘o’ stands for ‘omni’), released on May 13, 2024, by the giant OpenAI. This version extends the capabilities of ChatGPT-4. Based on user experiences, it has been concluded that ChatGPT-4o is faster and significantly better at generating human-like responses than ChatGPT-4. Numerous viral videos have showcased ChatGPT-4o's capacity to deliver responses in a chatty and occasionally flirtatious tone. Additionally, ChatGPT-4o features capabilities such as reading and discussing images, translating languages, identifying emotions from visual expressions, and remembering previous prompts. ChatGPT-4o has demonstrated new capabilities with the potential to revolutionize various fields, including education and e-learning [15].

## 4. LARGE LANGUAGE MODEL: DEFINITION, TYPES, AND COMPARISON

### 4.1. Definition

LLMs are specific implementations of transformer-based models trained on a massive amount of textual data, their foundation lies in their advanced processing capabilities and advanced algorithms. As a result, LLMs can capture, translate, expect, and produce textual content and other forms (images and voice notes). Also, they are flexible across diverse domains [2]. LLMs are known for achieving high performance in many natural language applications without task-specific training [10]. LLMs function simply by treating any NLP task as an oriented text generation task and generating output based on the input prompt, this last one describes the conditions that should be satisfied with the generated output [10].

### 4.2. Types of large language models

There are different types of LLMs, which can be categorized into five main groups. These groups include pre-trained and fine-tuned models, encoder-decoder models, multilingual models, domain-specific models, and autoregressive language models. This categorization is illustrated in Figure 3.

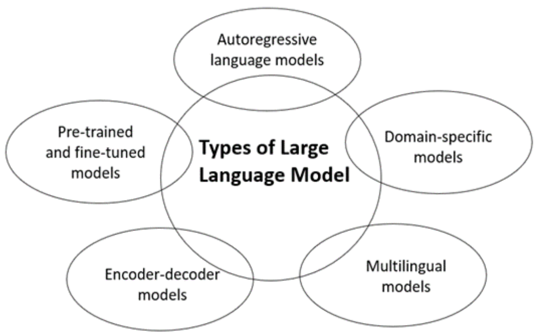


Figure 3. Categorization of LLMs based on architecture, application, and languages

#### 4.2.1. Pre-trained and fine-tuned models

Following the massive achievement of pre-trained image models in computer vision (VGGNet and AlexNet as examples) [16], NLP researchers got inspired to develop pre-trained models for NLP tasks, built based on transformers and self-supervised learning, and trained on a large corpus they can be tailored to fit goal-oriented tasks with constrained labeled data. The evolution of pre-trained language models (PLMs) begins with Google’s famous model BERT and OpenAI’s GPT-1 [9]. From there, many PLMs were developed, such as robustly optimized BERT pretraining approach (RoBERTa), XLNet, bidirectional and auto-regressive transformers (BART), and Electra [9].

#### 4.2.2. Encoder-decoder models

PLM's journey progressed in three stages: encoder-based models, decoder-based models, and encoder-decoder models. Valiyev *et al.* [17] explains the encoder-decoder architecture consists of both encoder and decoder modules; the encoder module is an embedding of encoder layers, and each encoder layer has self-attention and feed-forward networks. The decoder module is an embedding of decoder layers; each layer is equipped with a self-attention mechanism, masked self-attention, and feed-forward networks. Some encoder-decoder models are known for their high performance in natural language generation tasks, whereas others are known for performing well in both generation and language understanding tasks. MarianMT is an example of such a model besides text-to-text transfer transformer (T5), BART, and no language left behind (NLLB) [9].

#### 4.2.3. Multilingual models

The pre-trained LLMs have known a vast success in the English language, which inspired researchers to develop LLMs for other languages and multilingual LLMs. Some of the most widely known and used are multilingual bidirectional encoder representations from transformers (mBERT), cross-lingual language model (XLM), XLM-RoBERTa (XLM-R), multilingual text-to-text transfer transformer (mT5), multilingual decoding-enhanced BERT with disentangled attention (mDeBERTa), BigScience large open-science open-access multilingual language model (BLOOM), and BLOOM (instruction-tuned version) (BLOOMZ) [9], [10]. There are also famous bilingual LLMs, including Jais Arabic-Arabic (JAIS) LLM model (for English and Arabic) and general language model (GLM) (for English and Chinese) [10].

#### 4.2.4. Domain-specific models

Departing from low performance of general-domain LLMs in task and domain-specific problems, researchers shifted their focus to developing domain-specific models in various domains, for example, in education like ChatGPT. In healthcare, the Med-PaLM was designed to deliver top-tier answers to medical questions [9], while in finance, FinGPT and BloombergGPT emerged [10], and in law, SaulLM-7B LLM was tailored for the legal domain [18]. Additionally, in social media, social media pre-trained BERT (SoBert), a LLM trained on StackOverflow data for answering questions related to programming and software engineering [19], represents this specialized approach.

#### 4.2.5. Autoregressive language models

Auto-regressive LLMs are advanced models that can generate text by predicting the next token is created based on previously generated tokens. These models, such as autoregressive models, excel at tasks like language generation and completion by leveraging language patterns learned from training data. Examples of auto-regressive LLMs include GPT models developed by OpenAI; GPT-3, for example includes 175B parameters trained using a considerable corpus containing text from various resources such as web pages, Wikipedia, and books [11].

### 4.3. Traditional language models vs. large language models

Despite the effectiveness of traditional language models, they still are challenged by more complex language items such as idioms, metaphors, social context, and implicit emotions. However, LLMs show a marked increase in natural language understanding (NLU). They are trained with a variety of text resources to comprehend the beauty and complexity of human language. Table 1 shows how effectively LLMs improve performance in a variety of areas, but with added consideration for application relevance. After addressing the limitations of traditional language models across different application domains and how LLMs helped overcome them, Table 1 synthesizes the discussed ideas and provides an overview of their interrelations.

Table 1. The limitation of traditional language across various dimensions and LLM enhancements, a domain-based comparison

| Traditional language model limitation | LLM assistance | Relevant application domain |
|---|---|---|
| Limitation 1: performing sentiment analysis on textual content filled with complex emotions or sarcasm is challenging. | LLMs bridged this gap and helped machines with identifying and interpreting layered sentiments, ensuring high precision. | Market search [20], social media analytics [21], and customer feedback processing [22] |
| Limitation 2: in question-answering systems, the answer is generated according to keyword association and semantic reasoning. | LLMs in question answering systems allow them to capture the core of inquiries, taking into consideration various aspects, including context, in order to generate precise and authentic responses. | Customer support [22], academic research [23], and intelligent tutoring systems (ITS) [24] |
| Limitation 3: in content summarization, traditional LMs perform synthesis based on modest key sentences' extraction to an improved abstraction of the main idea. | LLMs have the ability to process massive corpora, understand the dominant themes, and generate to-the-point yet thorough outline summaries. | Academic research [23] and business intelligence [25] |

#### 4.3.1. Sentiment analysis

While traditional language models are used for sentiment analysis, there are certain issues with interpreting textual data in more complex emotion contexts, such as sarcasm [26]. LLMs address this issue, primarily through their ability to identify layered and complex sentiments. This is valuable in many fields, but the most relevant are market research, social media analysis, and processing customer feedback [27].

#### 4.3.2. Question answering systems

Traditional language models are applied in question-answering systems. The answers are generated based on given keyword association and semantic analysis [28]. Meanwhile, LLM-generated answers are built upon capturing the core of inquiries. LLMs take into consideration the context of a given question to create accurate and genuine answers. Question answering systems are applicable across diverse domains: customer support, academic search, and ITS, ensuring enhancements [29].

#### 4.3.3. Content summarization

In the context of content summarization, traditional language models' performance is weak [30]. LLMs perform synthesis based on modest key sentence extraction to generate an improved abstraction of the main idea. Concurrently, LLMs can process massive corpora and understand the dominant themes [31]. They are also capable of generating to-the-point yet thorough outline summaries. Relevant domains that apply to such a task are academic search, news aggregation, and business intelligence.

## 5. LARGE LANGUAGE MODEL: CHALLENGES AND LIMITATIONS

Though LLMs present various positive aspects and ensure specific outcomes, there are notable limitations inherent in their structure. One notable limitation is that it possesses significant computational demands. This aspect poses a challenge due to LLMs dependency on large training sets, which requires developing more efficient training algorithms. Solving that problem will protect the environment and reduce carbon footprints generated in creating and using LLMs [4], [32].

Additionally, data privacy and sensitivity present another constraint on the effectiveness of LLMs. This limitation arises from LLMs training on vast cross-sector amounts of data, which still raises a lot of concerns regarding unintentional disclosure of sensitive data by these models [32]. Moreover, these models are limited by biases ingrained in the data they were trained on. This factor results in the occurrence of "hallucinations", a situation where LLMs generate a response that is deceptively credible yet factually incorrect. Which creates the potential of building models that generate misleading information and "deepfakes" [4], [32].

Another limitation to be aware of correlates with the outcomes resulting from the rising adoption of LLMs integration into technical and non-technical sectors. No one denies it has valuable impact on automation, efficiency, and innovation. However, it has profoundly altered the economic and social landscape, particularly regarding job displacement within the workforce [32].

Furthermore, the model's interpretability introduces a significant consideration in applying LLMs. Deep learning models are known to be "black box" models with a vast number of parameters; hence, understanding and answering why the model opted for a specific decision can be hard to find [4], [32]. Concluding the list, the sixth limitation is related to the fact that LLMs lack human intuition, ethical reasoning, and accumulation of experiential knowledge. Despite their ability to generate responses that mimic human expertise, the users still need to educate themselves about the model's nature to adjust their level of trustworthiness and dependence on its output [4], [32].

## 6. LARGE LANGUAGE MODELS IMPACT ON EDUCATION

LLMs have a huge impact on diverse sectors; education is two of them. LLMs provide a more adaptive, appealing, and efficient learning environment for learners from different backgrounds and fields [33]. LLMs have become a transformative force in education, driving advancements across various educational functionalities. Educational technologies are generally categorized into four main functionalities [34]. In this part, we will elaborate on the use of LLMs in each of these functionalities with examples from real-world applications. Table 2 sums up the findings in terms of implementation of LLMs across different educational functionalities.

### 6.1. Personalization and adaptivity

Being one of the most important features in current e-learning systems, this is focused on customized adaptation of the process of learning and teaching based on individual students' skills as well as preferences [35]. LLMs can be very helpful in doing this by offering guided learning [35], dynamically adjusting content to the progress of a student [15], [33], [36]. Creating customized content according to learning styles, requirements, and performance information [33], [37]. Improving customized learning pathways, enabling students to proceed at their pace paths [15], [38], [39]. Designing adaptive tests, like computer-based multiple-choice questions (MCQs), in dependence upon learner profiles and information [40].

### 6.2. Assessment and evaluation

LLMs provide various means of making the assessment process more efficient and lightening the load of educators. Some examples of use are smart grading, where essays, short answers, or even code can be automatically graded for time and consistency savings [38], [41]. Automated test generation, like multimedia-supported tests tailored to target specific learning objectives [24], [42]. Automated feedback generation to provide instant, constructive feedback to students [24], [43], [44]. Discussion facilitation, through analysis and engagement in group discussions, to allow deeper peer-to-peer interaction [42], [45]. Last but not least, LLM-assisted classroom assistance, in which models assist both students and instructors during live sessions—for example, summarizing conversations, answering in-session queries, or giving explanations on demand [46].

### 6.3. Profiling and predicting

LLMs can analyze student behavior, activity, and performance so as to assist predictive analytics. Examples include identifying levels of comprehension by analyzing the submissions and interactions of students [3], [47]. Suggesting learning activities, tailored to current performance [3]. Sentiment analysis, so that learner frustration, confusion, or satisfaction can be identified, allowing the instructor to intervene preemptively [36], [43].

### 6.4. Intelligent tutoring systems

LLMs can facilitate ITS to implement more advanced, interactive features such as one-on-one virtual tutoring simulating human-like dialogue tailored to the student's learning requirements [3], [39], [48], [49]. Automated question-and-answer, where students may ask follow-up questions at any time [3], [49], [50]. Interactive content generation, e.g., learning games, scenarios, and simulations [24]. Enabling student discussions with prompts or intelligent agents to guide the discussion, and providing instant feedback, replicating a tutor's promptness during learning exercises [15], [45]. Designing practice exercises, e.g., personalized quizzes, logic puzzles, and role-playing scenarios [15], [45], [49], [51]. Mimicking tutor-like actors, e.g., chatbot-based tutors like Nemobot, which provide constant support for learning [52].

Table 2. Implementation of LLM applications across main educational functionalities

| Educational functionality | LLM application |
|---|---|
| Personalization and adaptivity | – Guided reading [35]<br>– Content generation [15], [33]<br>– Automated problem reframing across academic domains [33]<br>– Enhancing personalized learning paths [15], [38], [39]<br>– Automated creation of MCQs [40] |
| Evaluation and assessment | – Smart grading [38], [41]<br>– Generating various resources, including assessments and multimedia content [24], [42]<br>– Providing automated, process-oriented feedback to students [24], [43]<br>– Support group conversations [42], [45]<br>– automated creation of MCQs [40]<br>– LLM-driven classroom flipping [46] |
| Profiling and predicting | – Identifying student's comprehension level [3], [47]<br>– Determine suitable tasks for each learner [3]<br>– Sentiment analysis [43] |
| ITS | – Providing one-on-one tutoring for students [3], [39]<br>– Answering questions [3], [50]<br>– Generating interactive content, facilitating discussions, and providing real-time feedback [24]<br>– Real-world practice through generating personalized puzzles, quizzes, and role-playing scenarios [15], [45]<br>– Feedback generation [53]<br>– Simulate a tutor-like experience with a chatbot like Nemobot [52] |

## 7. CHALLENGES AND FUTURE DIRECTIONS

Recent studies have shown a growing number of researchers exploring the application of LLMs in education for tasks like academic research, assessing knowledge, course planning, and more. However, many challenges and opportunities remain to be addressed. Having addressed each challenge separately, it's useful to provide a clearer view. Figure 4 summarizes the identified challenges.

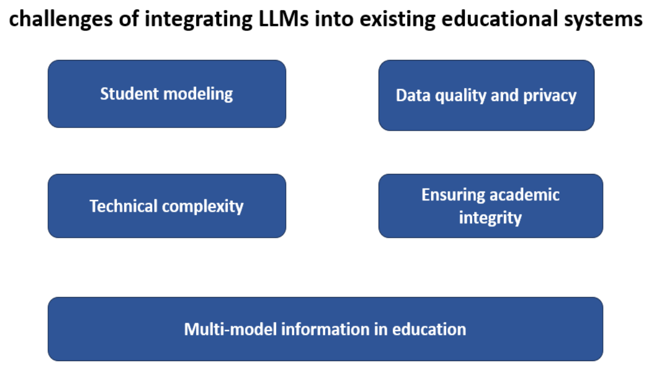


Figure 4. Key challenges in integrating LLMs within educational contexts and e-learning systems

### 7.1. Student modeling

Before the advancement of LLMs, student behavior modeling in the deep learning era relied on sequential models like RNNs and transformers, which lacked direct student feedback and often resulted in outcomes that were difficult to interpret. The introduction of LLM-based educational systems enables students to express their needs through dialogue, allowing for the extraction of personalized characteristics such as topic mastery and learning preferences. Moreover, LLMs' abilities to simulate human behavior and generate human-like samples offer promising potential for enhancing student simulations. For students with limited interaction records, LLM-based simulators can generate additional data, providing valuable insights for better understanding individual learners and improving educators' teaching strategies [51].

### 7.2. Data quality and privacy

Privacy of data and privacy of personal information is one of the top concerns with using LLMs because using LLMs will require collecting data on learners and teachers personal and learning data, therefore creating privacy protection concerns around data collection. Schools and programs must protect and ensure the privacy of their students/learners and educators while protecting and securing the data that is collected. Parents and educators are also critical for helping educate children on data privacy and safety and managing risks when using LLMs. In regard to the collection and use of student learning data there have to be appropriate privacy protections, this will help ensure that student learning data is not hacked and that data is not used for unintended outcomes. As the world is entering an unusual period of intelligence-assisted education, seen through LLM models like ChatGPT, GPT-4, and MathGPT, LLMs will take education beyond limitations and we need to continue to find methods to "train" LLMs more effectively and make improvements to large-scale models as a whole [54].

### 7.3. Technical complexity

LLMs are becoming increasingly complex, with computation times for single-step processes increasing more than tenfold. Tasks that once took a few hours to train now take several days, with the need to maintain a one-day limit for testing still essential. The cost of training large, general-purpose models is significant and escalates when optimization, updates, and deployment are included. Integrating LLMs into existing education systems presents significant technical challenges. Schools may struggle to align LLM capabilities with their current IT infrastructures, which may require significant upgrades and customization to support these advanced technologies. In addition, ChatGPT's reliance on a large number of GPU chips for data processing adds further complexity and cost to deployment [54].

### 7.4. Ensuring academic integrity

LLMs produce text that can rival or even surpass human writing in terms of fluency and quality. While this article highlights their potential in educational contexts, it is important to recognize that over-reliance on LLMs can inhibit natural learning processes. Using LLMs for homework help is acceptable, but relying on them to complete tasks can limit opportunities for meaningful practice and understanding. Detecting LLM-generated content is essential to prevent cheating and maintain educational integrity. Recent studies have introduced two main detection strategies: statistical outlier detection, which identifies variations in linguistic features, and supervised classifiers, which use machine learning algorithms trained on examples of both human and LLM-generated texts. As LLMs evolve, detection methods must also evolve, incorporating statistical analysis, machine learning, and potentially new techniques to effectively distinguish LLM output from human-written content [51].

### 7.5. Multi-modal information in education

Multimodal information in education is not a novel phenomenon. For example, geometric word problems include images along with text and textbooks often utilize text in conjunction with images for a concept too. This representation and multimodal alignment are part of what makes an intelligent education system optimal, and ultimately what makes a multimodal information system more robust. The multimodal landscape is rapidly evolving, especially for LLMs, as various architectures and pretraining tasks are being established as well. However, educational contexts represent a unique and complex case to deal with multimodal information. Images and text in the educational context are often very precisely aligned with one another. For example, geometry word problems can have very specific information about the parameters of the geometric shapes in the images as well as aesthetic details, that require equally precise verbal representation as well. The model has to be able to represent the image all the information and details conveyed in the visual modality. Cross-model reasoning is typically required for educational multimodal information representation and datasets do not usually represent cross-modal reasoning complexities.

## 8. CONCLUSION

This review goes beyond high-level promises and makes specific statements of how LLMs can be a force for better e-learning, and what needs to be managed, in order to do so responsibly. First, our education-focused taxonomy makes an explicit connection between LLM capabilities and four pedagogical functions, which can usefully inform selection and evaluative efforts. Second, our domain-based comparison clarifies which specific areas LLMs have unambiguous advantages over traditional pedagogical models (i.e., nuanced sentiment, contextual question answering, and true abstractive summarization) and how that might mean better feedback, guidance, or study support for learners. Third, we aggregate six systemic threats to their use (computational cost/carbon, privacy and data governance, bias and hallucination, workforce challenges, model opacity, and limits of human judgment) as design opportunities (to design for privacy-by-design data management, bias audits, verifiable grounding, educator-in-the-loop oversight, and explainability for pedagogical practices). Finally, our reflexive methodology for documenting our review process and curated body of documents creates a living resource for recaps, or preparing for advances in more substantive model development, policy regulations, or classroom practice.


## ACKNOWLEDGMENTS

The authors would like to thank the Artificial Intelligence and Complex Systems Engineering Laboratory (ENSAM, Hassan II University) and the Smartilab Laboratory (Moroccan School of Engineering) for their support.

## FUNDING INFORMATION

This research was supported by the Artificial Intelligence and Complex Systems Engineering Laboratory (ENSAM, Hassan II University, Casablanca, Morocco) and Laboratory Smartilab (Moroccan School of Engineering Sciences, Rabat, Morocco)


## AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

| Name of Author | C | M | So | Va | Fo | I | R | D | O | E | Vi | Su | P | Fu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Laila Boullous | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | |
| Mustapha Hain | | | | ✓ | | ✓ | | ✓ | | ✓ | ✓ | ✓ | | |
| Adil Chergui | ✓ | ✓ | | ✓ | | ✓ | | | | ✓ | | ✓ | | |
| Brahim Elbhiri | | ✓ | | | | ✓ | | | | ✓ | | ✓ | | |

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

## CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

## INFORMED CONSENT

Not applicable. This study did not involve human participants, human data, or any personally identifiable information. All data used were either publicly available, fully anonymized, or derived from non-human sources, and therefore no informed consent was required from individuals.

## ETHICAL APPROVAL

Not applicable. This research did not involve human subjects, human biological materials, or experimental procedures on animals. The work was conducted solely on computational models, publicly available datasets, or non-sensitive data that did not require intervention with living organisms. Therefore, ethical approval from an institutional review board or animal ethics committee was not necessary for this study.

## DATA AVAILABILITY

As this is a review article, all data used in this study were retrieved from established scientific databases and academic articles. The comprehensive list of references provided in this manuscript contains all the necessary information and sources utilized for the analysis. Therefore, data availability is not applicable to this paper as no new data were created or analyzed in this study.

## REFERENCES


[1] A. Vaswani *et al.*, "Attention is all you need," in *31st Conference on Neural Information Processing Systems (NIPS 2017)*, 2017, pp. 1–11.

[2] S. Chatterjee, M. Bhattacharya, S. Pal, S. Lee, and C. Chakraborty, "ChatGPT and large language models in orthopedics: from education and surgery to research," *Journal of Experimental Orthopaedics*, vol. 10, no. 1, Jan. 2023, doi: 10.1186/s40634-023-00700-1.

[3] Y. K. Dwivedi *et al.*, "Opinion paper: 'So what if ChatGPT wrote it?' Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy," *International Journal of Information Management*, vol. 71, Aug. 2023, doi: 10.1016/j.ijinfomgt.2023.102642.

[4] E. Y. Zhang, A. D. Cheok, Z. Pan, J. Cai, and Y. Yan, "From turing to transformers: a comprehensive review and tutorial on the evolution and applications of generative transformer models," *Sci*, vol. 5, no. 4, Dec. 2023, doi: 10.3390/sci5040046.

[5] D. Qiu and B. Yang, "Text summarization based on multi-head self-attention mechanism and pointer network," *Complex & Intelligent Systems*, vol. 8, no. 1, pp. 555–567, Feb. 2022, doi: 10.1007/s40747-021-00527-2.

[6] Z. Ahmed, S. S. Shanto, and A. I. Jony, "Potentiality of generative AI tools in higher education: evaluating ChatGPT's viability as a teaching assistant for introductory programming courses," *STEM Education*, vol. 4, no. 3, pp. 165–182, 2024, doi: 10.3934/steme.2024011.

[7] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North*, 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.

[8] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," *OpenAI*, 2018.

[9] W. Ali and S. Pyysalo, "A survey of large language models for European languages," Aug. 2024, *arXiv:* 2408.15040.

[10] K. S. Kalyan, "A survey of GPT-3 family large language models including ChatGPT and GPT-4," *Natural Language Processing Journal*, vol. 6, Mar. 2024, doi: 10.1016/j.nlp.2023.100048.

[11] H. Fan, B. Li, Y. Weng, and S. Zhou, "Small but mighty: enhancing time series forecasting with lightweight LLMs," *The Journal of Supercomputing*, vol. 81, no. 8, Jun. 2025, doi: 10.1007/s11227-025-07491-5.

[12] K. Zhou, M. Constantinides, and D. Quercia, "Should LLMs be WEIRD? exploring WEIRDness and human rightsin large language models," in *Proceedings of the Eighth AAAI/ACM Conference on AI, Ethics, and Society (AIES2025)*, 2025, pp. 2808–2820.

[13] H. Yang, X.-Y. Liu, and C. Dan Wang, "FinGPT: open-source financial large language models," *SSRN Electronic Journal*, 2023, doi: 10.2139/ssrn.4489826.

[14] K. Singhal *et al.*, "Toward expert-level medical question answering with large language models," *Nature Medicine*, vol. 31, no. 3, pp. 943–950, Mar. 2025, doi: 10.1038/s41591-024-03423-7.

[15] S. Pang, E. Nol, and K. Heng, "ChatGPT-4o for English language teaching and learning: features, applications, and future prospects," *SSRN Electronic Journal*, 2024, doi: 10.2139/ssrn.4837988.

[16] E. Puls, M. Todescato, and J. Carbonera, "An evaluation of pre-trained models for feature extraction in image classification," in *Proceedings of the 26th International Conference on Enterprise Information Systems*, 2024, pp. 465–476, doi: 10.5220/0012622300003690.

[17] G. Valiyev, P. Eles, A. Kok, and R. D'Ercole, "Comparison of transformer models for performance on domain specific texts: a systematic evaluation of intrinsic model performance," in *2023 International Conference on Military Communications and Information Systems (ICMCIS)*, May 2023, pp. 1–6, doi: 10.1109/ICMCIS59922.2023.10253491.

[18] T. V. Q. Le, N. P. Nguyen, T.-C. Duong, A.-C. Le, V.-C. Nguyen, and V.-H. Nguyen, "Enhancing reading comprehension of Vietnamese LLMs with synthetic data," in *2024 16th International Conference on Knowledge and System Engineering (KSE)*, Nov. 2024, pp. 327–332, doi: 10.1109/KSE63888.2024.11063523.

[19] W. Alhoshan, A. Ferrari, and L. Zhao, "Zero-shot learning for requirements classification: an exploratory study," *Information and Software Technology*, vol. 159, Jul. 2023, doi: 10.1016/j.infsof.2023.107202.
[20] J. Brand, A. Israeli, and D. Ngwe, "Using LLMs for market research," *Working Paper 23-062*. Harvard Business School, Boston, MA, United States, 2023.
[21] H. Guo *et al.*, "SNS-Bench: defining, building, and assessing capabilities of large language models in social networking services," *Proceedings of the 42nd International Conference on Machine Learning*, 2025, pp. 1-37.
[22] S. Gunathilaka and N. de Silva, "Automatic analysis of app reviews using LLMs," in *Proceedings of the 17th International Conference on Agents and Artificial Intelligence*, 2025, pp. 828–839, doi: 10.5220/0013375600003890.
[23] T. Mishra *et al.*, "Use of large language models as artificial intelligence tools in academic research and publishing among global clinical researchers," *Scientific Reports*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-81370-6.
[24] T. Nguyen, "Leveraging large language models for educational enhancement: a case study of ChatGPT, BingChat, and Bard," *Preprints*, Sep. 22, 2023, doi: 10.20944/preprints202309.1554.v1.
[25] L. Weng *et al.*, "DataLab: a unified platform for LLM-powered business intelligence," in *2025 IEEE 41st International Conference on Data Engineering (ICDE)*, May 2025, pp. 4346–4359, doi: 10.1109/ICDE65448.2025.00326.
[26] X. Pu, G. Wu, and C. Yuan, "Exploring overall opinions for document level sentiment classification with structural SVM," *Multimedia Systems*, vol. 25, no. 1, pp. 21–33, Feb. 2019, doi: 10.1007/s00530-017-0550-0.
[27] M. Kumar, L. Khan, and H.-T. Chang, "Evolving techniques in sentiment analysis: a comprehensive review," *PeerJ Computer Science*, vol. 11, Jan. 2025, doi: 10.7717/peerj-cs.2592.
[28] B. Ojokoh and E. Adebisi, "A review of question answering systems," *Journal of Web Engineering*, vol. 17, no. 8, pp. 717–758, 2019, doi: 10.13052/jwe1540-9589.1785.
[29] T. Huynh and D. Nguyen, "Collaborating agents with large language models for question-answering tasks," in *2025 IEEE 5th International Conference on Software Engineering and Artificial Intelligence (SEAI)*, Jun. 2025, pp. 1–5, doi: 10.1109/SEAI65851.2025.11108177.
[30] A. P. Widyassari *et al.*, "Review of automatic text summarization techniques & methods," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 4, pp. 1029–1046, Apr. 2022, doi: 10.1016/j.jksuci.2020.05.006.
[31] W. M. Aly, T. H. A. Soliman, and A. M. AbdelAziz, "Cross-domain evaluation of large language models for abstractive text summarization: an empirical perspective," *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 6, 2025.
[32] D.-M. Petroșanu, A. Pîrjan, and A. Tăbușcă, "Tracing the influence of large language models across the most impactful scientific works," *Electronics*, vol. 12, no. 24, Dec. 2023, doi: 10.3390/electronics12244957.
[33] H. Einarsson, S. H. Lund, and A. H. Jónsdóttir, "Application of ChatGPT for automated problem reframing across academic domains," *Computers and Education: Artificial Intelligence*, vol. 6, Jun. 2024, doi: 10.1016/j.caeai.2023.100194.
[34] L. Boullous, M. Hain, A. Chergui, and B. E. Bhiri, "AI's contribution in e-learning: a review and suggestion for hybridizing existing classification approaches," in *Technology and the Environment: Implementing Smart and Sustainable Solutions into Our Cities*, 2025, pp. 11–22, doi: 10.1007/978-3-031-74474-7_2.
[35] J. Gao and S. Li, "Can students use ChatGPT for academic purposes? the potential and challenges of using large language models in post-secondary education," *Antistasis*, vol. 14, no. 1, pp. 144–157, 2025.
[36] Q. Shi, C. E. Jimenez, S. Yao, N. Haber, D. Yang, and K. Narasimhan, "When models know more than they can explain: quantifying knowledge transfer in human-AI collaboration," Jun. 2025, *arXiv: 2506.05579.*
[37] W. Gan, Z. Qi, J. Wu, and J. C.-W. Lin, "Large language models in education: vision and opportunities," in *2023 IEEE International Conference on Big Data (BigData)*, Dec. 2023, pp. 4776–4785, doi: 10.1109/BigData59044.2023.10386291.
[38] Z. H. İpek, A. İ. C. Gözüm, S. Papadakis, and M. Kallogiannakis, "Educational applications of the ChatGPT AI system: a systematic review research," *Educational Process International Journal*, vol. 12, no. 3, 2023, doi: 10.22521/edupij.2023.123.2.
[39] T. K. Nepal, "Exploring the applications and challenges of ChatGPT in research and academia: a comprehensive review," *West Science Interdisciplinary Studies*, vol. 2, no. 5, pp. 1043–1050, May 2024, doi: 10.58812/wsis.v2i05.729.
[40] C. N. Hang, C. Wei Tan, and P.-D. Yu, "MCQGen: a large language model-driven MCQ generator for personalized learning," *IEEE Access*, vol. 12, pp. 102261–102273, 2024, doi: 10.1109/ACCESS.2024.3420709.
[41] S. Tobler, "Smart grading: a generative AI-based tool for knowledge-grounded answer evaluation in educational assessments," *MethodsX*, vol. 12, Jun. 2024, doi: 10.1016/j.mex.2023.102531.
[42] B. Selvanathan and S. Narayanan, "ChatGPT in higher education Malaysia: an opportunity or threat to the education system?," *International Journal of Academic Research in Progressive Education and Development*, vol. 13, no. 3, Jun. 2024, doi: 10.6007/IJARPED/v13-i3/21455.
[43] S. Yang, Y. Dong, and Z. G. Yu, "ChatGPT in education," *International Journal of Information and Communication Technology Education*, vol. 20, no. 1, pp. 1–19, May 2024, doi: 10.4018/IJICTE.346826.
[44] J. Stamper, R. Xiao, and X. Hou, "Enhancing LLM-based feedback: insights from intelligent tutoring systems and the learning sciences," in *25th International Conference on Artificial Intelligence in Education (AIED 2024) BlueSky special track*, 2024, pp. 32–43, doi: 10.1007/978-3-031-64315-6_3.
[45] J. C. Cronjé, B. C. Swartz, and L. Z. Valentine, "Assessing the potential of chatbots as tutoring tools in engineering education: a South African case study," in *4th International Conference on Teaching, Assessment and Learning in the Digital Age*, 2023, pp. 1–13.
[46] Y. Ling, "AI-enabled learning analytics and intelligent interaction system for Chinese college English flipped classrooms: model reconstruction and effectiveness evaluation," in *Proceedings of the 2026 5th International Conference on Big Data, Information and Computer Network*, Jan. 2026, pp. 103–109, doi: 10.1145/3801228.3801244.
[47] S. Alan and E. Yurt, "Flipped learning: an innovative model for enhancing education through ChatGPT," *International Journal of Modern Education Studies*, vol. 8, no. 1, Jun. 2024, doi: 10.51383/ijonmes.2024.328.
[48] Sneha, V. Tyagi, S. Agrawal, V. Manjhi, and A. Sharma, "Empowering educators and engaging learners: a framework for a blended learning management system," in *2025 IEEE 7th International Conference on Computing, Communication and Automation (ICCCA)*, Nov. 2025, pp. 1–5, doi: 10.1109/ICCCA66364.2025.11325623.
[49] Y. Chen, N. Ding, H.-T. Zheng, Z. Liu, M. Sun, and B. Zhou, "Empowering private tutoring by chaining large language models," in *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, Oct. 2024, pp. 354–364, doi: 10.1145/3627673.3679665.
[50] T. K. Nepal, "The role of ChatGPT in exploring science and its applications: a case study at the 'Citizen' program," *Asian Journal of Research in Computer Science*, vol. 17, no. 5, pp. 167–171, Mar. 2024, doi: 10.9734/ajrcos/2024/v17i5446.
[51] M. A. Razafinirina, W. G. Dimbisoa, and T. Mahatody, "Pedagogical alignment of large language models (LLM) for personalized learning: a survey, trends and challenges," *Journal of Intelligent Learning Systems and Applications*, vol. 16, no. 4, pp. 448–480, 2024, doi: 10.4236/jilsa.2024.164023.

[52] Y. Wang, S. Guo, L. Ling, and C. W. Tan, "Nemobot: crafting strategic gaming LLM agents for K-12 AI education," in *Proceedings of the Eleventh ACM Conference on Learning @ Scale*, Jul. 2024, pp. 393–397, doi: 10.1145/3657604.3664671.

[53] T. Ruwe and E. M.-Paus, "Embracing LLM feedback: the role of feedback providers and provider information for feedback effectiveness," *Frontiers in Education*, vol. 9, Oct. 2024, doi: 10.3389/feduc.2024.1461362.

[54] S. Zhao, Y. Tan, and J. Zheng, "Application and development prospects of large language models in the field of education: a survey," in *2025 14th International Conference on Educational and Information Technology (ICEIT)*, Mar. 2025, pp. 45–51, doi: 10.1109/ICEIT64364.2025.10975965.


## BIOGRAPHIES OF AUTHORS

**Laila Boullous** is a Ph.D. student at ENSAM Casablanca, Hassan II University, affiliated with the Artificial Intelligence and Complex Systems Engineering (AICSE) Laboratory. As a passionate teacher and researcher, her work focuses primarily on the application of artificial intelligence in e-learning, exploring innovative ways to enhance education through technology. She can be contacted at email: Laila.boullous-etu@etu.univh2c.ma.

**Mustapha Hain** is a Ph.D. researcher and faculty member at ENSAM Casablanca, part of Hassan II University, and is currently the director of the Artificial Intelligence and Complex Systems Engineering (AICSE) Laboratory. He is also a visiting researcher for arts et métiers. He's research focus is in the areas of digital systems, e-logistics, software modeling, and supply chain management. As an active member of the scientific community, he serves as a board member and guest reviewer for many journals and conferences. He can be contacted at email: infohain@gmail.com.

**Adil Chergui** serves as a university professor and Ph.D. researcher at ENSAM Casablanca, Hassan II University, and is a working member of the Artificial Intelligence and Complex Systems Engineering (AICSE) Laboratory. His research interests are broad and include machine learning, data science, neural networks, and artificial intelligence with applications including, but not limited to, image and signal processing, pattern recognition, and computer vision. His expertise greatly contributes in the area of projects pertaining to feature extraction, digital image processing, and machine intelligence. He can be contacted at email: adil.chergui@ensam-casa.ma.

**Brahim Elbhiri** is a researcher and expert in embedded systems at EMSI. He works in the areas of computer communications, computer graphics, and artificial neural networks, particularly on image and signal processing, feature extraction, and efficient clustering in heterogeneous wireless sensor networks. He also has experience with various applications in digital and video signal processing. He can be contacted at email: B.Elbhiri@emsi.ma.