Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection School Of Computing and Information Systems | School of Computing and Information Systems

1-2023

Seven pillars for the future of Artificial Intelligence

Erik CAMBRIA

Rui MAO

Melvin CHEN

Zhaoxia WANG
Singapore Management University, zxwang@smu.edu.sg

Seng-Beng HO

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research

Part of the Artificial Intelligence and Robotics Commons

Citation

CAMBRIA, Erik; MAO, Rui; CHEN, Melvin; WANG, Zhaoxia; and HO, Seng-Beng. Seven pillars for the future of Artificial Intelligence. (2023). *IEEE Intelligent Systems*. 62-69.
Available at: https://ink.library.smu.edu.sg/sis_research/8319

This Journal Article is brought to you for free and open access by the School of Computing and Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Computing and Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylds@smu.edu.sg.

Seven Pillars for the Future of Artificial Intelligence

Erik Cambria, Rui Mao, Melvin Chen
Nanyang Technological University, Singapore

Zhaoxia Wang
Singapore Management University, Singapore

Seng-Beng Ho
Agency for Science, Technology and Research, Singapore

***Abstract*—In recent years, AI research has showcased tremendous potential to impact positively humanity and society. Although AI frequently outperforms humans in tasks related to classification and pattern recognition, it continues to face challenges when dealing with complex tasks such as intuitive decision-making, sense disambiguation, sarcasm detection, and narrative understanding, as these require advanced kinds of reasoning, e.g., commonsense reasoning and causal reasoning, which have not been emulated satisfactorily yet. To address these shortcomings, we propose seven pillars that we believe represent the key hallmark features for the future of AI, namely: *Multidisciplinarity*, *Task Decomposition*, *Parallel Analogy*, *Symbol Grounding*, *Similarity Measure*, *Intention Awareness*, and *Trustworthiness*.**

IN 2022, the world was stunned by ChatGPT, a chatbot that relies on a large language model (LLM) built by means of generative pre-training transformers (GPT). We do not deny the performance capabilities of GPT-based LLMs: these capabilities enable chatbots to generate detailed, original, and plausible responses to prompts. GPT-4 and other LLMs are pretrained on a large dataset (self-supervised and at scale), before being adapted for a variety of downstream tasks through fine-tuning. Pre-training is time-intensive and never repeated, whereas fine-tuning is conducted in a regular fashion.

The behavior of GPT-based chatbots such as ChatGPT and ChatGPT Plus arises through fine-tuning. The performance capabilities of LLMs have been attributed to at least two factors: pre-training and scale [1]. Pretraining, an instance of transfer learning in which LLMs use knowledge acquired from one task (source) and transfer this knowledge to another task (target), makes LLMs possible. Scale, including better computer hardware, the transformer architecture, the availability of more and higher-quality training data, makes LLMs powerful. Although these capabilities are not insubstantial, they do not yet rise to the level of natural language understanding [2], [3], [4].

 © 2023 AI Focus

In addition, LLMs are prone to hallucination: ChatGPT may produce linguistic responses that, though syntactically and semantically fine and credible-sounding, are ultimately incorrect [5]. Furthermore, we may distinguish between the capabilities of LLMs (acquired through pretraining) and the behavior (affected by fine-tuning, which happens after pretraining) of LLMs. Fine-tuning can have unintended effects, including behavioral drift on certain tasks. As discussed in a recent study [6], in fact, ChatGPT seems prone to the 'short blanket dilemma': while trying to improve its accuracy on some tasks, OpenAI researchers inadvertently made ChatGPT worse for tasks which it previously excelled at.

AI research has slowly been drifting away from what its forefathers envisioned back in the 1960s. Instead of evolving towards the emulation of human intelligence, AI research has regressed into the mimicking of intelligent behavior in the past decade or so. The main goal of most tech companies is not designing the building blocks of intelligence but simply creating products that existing and potential customers deem intelligent. In this context, instead of labeling it as 'artificial' intelligence, it may be more apt to characterize such research as 'pareidoliac' intelligence. This term highlights the development of expert systems while raising questions about their claim to possess genuine intelligence.

We feel there is a need for an AI refocus on humanity, an Anti-Copernican revolution of sorts: like Copernicus demoted humans from their privileged spot at the center of the universe, in fact, deep learning has removed humans from the equation of learning. In traditional neural networks, especially those with a shallow architecture (few hidden layers), humans were at the center of the technological universe as they had to carefully design the input features, select appropriate hyperparameters, adjust learning rates, etc. Instead, due to their increased complexity and capacity to automatically learn features from data, deep neural networks do not require manual feature engineering and, hence, have effectively removed humans from the loop of learning. While this is good in terms of cost, time, and effectiveness, it is bad for several other reasons, including transparency, accountability, and bias.

In the deep learning era, humans no longer have control on how the learning process takes place. To save on cost and time, we delegated the important task of selecting which features are important for classification to deep neural networks. These, however, are mathematical models with no commonsense whatsoever: they do not know how to properly choose features. For example, in selecting candidates for a job opening, deep neural networks may decide that gender is an important feature to take into account simply because more men are present in the training data as positive samples.

The issue is not only that deep nets may accidentally choose unimportant or even wrong features, but that we have no way of knowing this because of their black-box nature. In other words, not only humans have been taken out of the picture but they have also been blindfolded. For these reasons, we feel there is a need to bring human-centered capabilities back at the center of AI, e.g., by having human-in-the-loop or human-in-command systems that ensure AI outputs and reasoning steps are human-readable and human-editable. To this end, we propose seven pillars for the future of AI (Fig. 1), namely: *Multidisciplinarity* (Section I), *Task Decomposition* (Section II), *Parallel Analogy* (Section III), *Symbol Grounding* (Section IV), *Similarity Measure* (Section V), *Intention Awareness* (Section VI), and *Trustworthiness* (Section VII). The focus of our 'manifesto' is on natural language processing (NLP) but the same concepts can be easily adapted to other AI domains such as computer vision, speech recognition, signal processing, multimodal analysis, edge computing, and robotics.

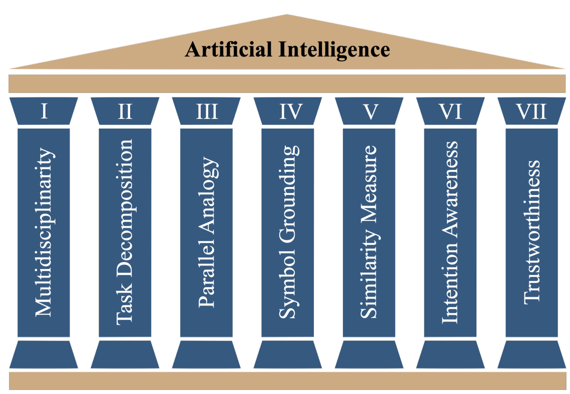

Figure 1. Seven Pillars for the future of AI.

I. Multidisciplinarity

Due to the complex and multifaceted nature of modern AI technologies and applications, *Multidisciplinarity* is of increasing importance for the future of AI. The integration of knowledge from disciplines like mathematics, semiotics, logic, linguistics, psychology, sociology, and ethics allows for a more holistic understanding of AI's capabilities and limitations. Mathematical principles such as linear algebra, calculus, probability theory, and optimization underpin the design of AI algorithms. Maths alone, however, is not enough for designing intelligent systems, because mathematical approaches excel at capturing predominant linguistic patterns but often struggle with addressing 'long tail' issues such as less common or niche linguistic phenomena.

Disciplines like semiotics can help AI systems understand the nuances of language, including metaphors, idioms, sarcasm, and cultural references, whether they fall within the more frequent or rarer occurrences across the spectrum of everyday human language. Logic also plays a fundamental and enduring role in the development and advancement of AI, as it provides a rigorous framework for reasoning, problem-solving, and knowledge representation. Word embeddings, which essentially replace words with numbers, have made most AI researchers forget about the importance of linguistics. Concepts from syntax, semantics, phonetics, and morphology (see Section II), however, are crucial for interpreting the intended meaning of natural language.

Psychology will play an essential role in creating systems that enhance well-being, foster human relationships, and provide meaningful and empathetic interactions. By addressing issues related to inequality and cultural diversity, sociology will guide AI development in ways that promote positive societal outcomes and responsible innovation. The arts are also going to be key for the future of AI, as highlighted by recent STEAM (STEM + Art) initiatives, in order to 'humanize' AI through computational creativity, cultural and social understanding, and the enhancement of AI usability [7]. Finally, ethics are paramount to ensure that AI technologies are developed, deployed, and used in ways that align with human values and promote accountability [8].

II. Task Decomposition

Like *Multidisciplinarity*, *Task Decomposition* aims to better handle the complex and multifaceted nature of AI problems. It is a method commonly used in psychology, instructional design, and project management to break down a complex task or activity into its individual components. *Task Decomposition* is also important for NLP: no matter what kind of downstream task we are handling, if we do not deconstruct it into its constituent subtasks, we are practically forcing our model to implicitly solve a lot of subtasks it has never been trained for. The 'sentiment suitcase model' [9], for example, lists 15 NLP subtasks that need to be solved separately before sentiment analysis can be accomplished (Fig. 2).

Firstly, a Syntactics Layer pre-processes text so that informal and inflected expressions are reduced to plain standard text [10]. This is done through subtasks such as microtext normalization, for refining and standardizing informal text, part-of-speech (POS) tagging, for assigning grammatical categories (such as nouns, verbs, adjectives, and adverbs) to each word in a sentence, and lemmatization, for reducing words to their base or dictionary form (lemmas).

Secondly, a Semantics Layer deconstructs normalized text into concepts, resolves references, and filters out neutral content from the input [11]. This is done through subtasks such as word sense disambiguation, for determining the correct meaning of a word within a given context, named entity recognition, for identifying and classifying

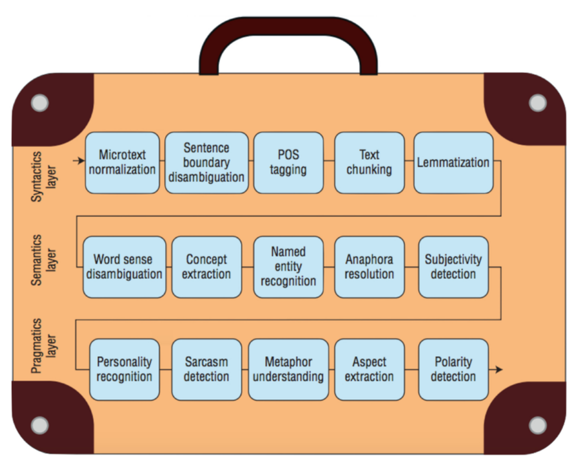

Figure 2. The sentiment suitcase model [9].

names of people, places, organizations, and dates, and subjectivity detection, to distinguish between factual information and subjective content.

Finally, the Pragmatics Layer extracts meaning from both sentence structure and semantics obtained from the previous layers. This is done through subtasks such as personality recognition, to infer traits, characteristics, preferences, and behavioral tendencies of the speaker, metaphor understanding, for interpreting figurative language in text, and aspect extraction, for identifying and extracting specific facets, features, or components mentioned in text and, hence, enabling a more fine-grained analysis. Only after handling all these subtasks, which we humans take care of almost subconsciously during reading or communication, the downstream task, e.g., polarity detection, can be effectively processed.

III. Parallel Analogy

Similar to *Multidisciplinarity* and *Task Decomposition*, *Parallel Analogy* looks at AI problems in a multifaceted way. Engineers and computer scientists have always been obsessed with optimization. In the development of AI systems, this translates into finding the 'best' knowledge representation, the 'best' reasoning algorithm, the 'best' way of doing things. This, however, results in only having one way of solving a problem. Instead, several analogous representations of the same problem should be maintained in parallel while trying to solve it so that, when problem-solving begins to fail while using one representation, the system can switch to one of the others.

Parallel Analogy, or 'panalogy' like the late Marvin Minsky used to call it [12], is key to solving highly complex AI problems, but also simpler problems in which a change of perspective is required. In affective computing tasks, for example, sometimes it is useful to see emotion concepts from a semantic point of view, e.g., 'joy' and 'sadness' are similar because they are both emotions, or a polarity point of view, e.g., 'joy' and 'sadness' are opposite because the former is positive and the latter is negative [13]. Similarly, we could say that words like 'joyful', 'joyfully', 'enjoy', and 'enjoyment' are similar because they all share the same root word 'joy', but totally different in a POS tagging sense (adjective versus adverb versus verb versus noun, respectively).

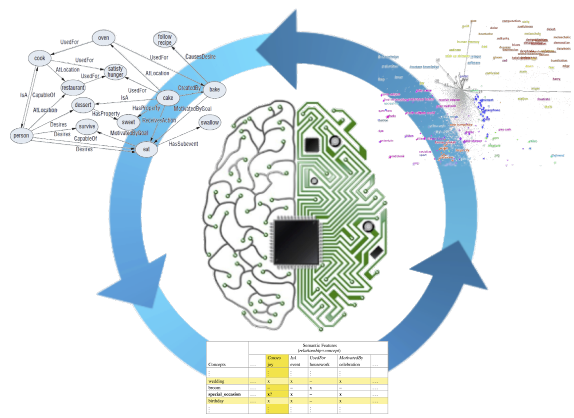

Figure 3. An example of 'panalogy' where the same data is 'redundantly' represented as a knowledge graph, as a matrix, and as embeddings [13].

For more general NLP tasks, it could be useful to have the same data 'redundantly' represented both as a knowledge graph and as embeddings (Fig. 3). The knowledge graph could be more useful for solving problems requiring *Symbol Grounding*, e.g., answering questions like 'what is what?' (see Section IV). Embeddings, instead, could be more useful for *Similarity Measure*, e.g., answering questions like 'what is similar to what?' (see Section V).

IV. Symbol Grounding

Symbol Grounding is the central pillar of our structure, being one of the fundamental challenges in the field of AI since its inception. It deals with how symbols, which are abstract representations, acquire meaning and connection to the real world. In human cognition, we understand the meanings of symbols through a process known as 'grounding'. When we see or hear a word, for example, our brains associate it with the sensory experiences and interactions we have had with the objects or concepts that such a word represents, thus providing a foundation for our understanding of its meaning.

In the context of AI, the *Symbol Grounding* problem arises because computers lack the inherent sensory experiences that humans possess. They process symbols as strings of characters or digital information without a direct connection to the real world, raising the question of how they can truly understand the meanings of symbols in a way that is equivalent to human understanding.

For instance, consider the word ‘apple’. Humans understand this word not just as a sequence of letters, but as a fruit with certain sensory qualities like color, taste, texture, and smell, all of which are grounded in our experiences with actual apples. Current AI systems are unable to grasp the richness of meaning behind the word ‘apple’ without having those sensory experiences. To solve this, we may have to take a step back in order to move forward. Old-school (symbolic) AI was better at *Symbol Grounding* but it was not scalable nor flexible. New deep-learning-based (subsymbolic) AI, instead, is very scalable and flexible but it does not handle symbols. The best of both worlds could be achieved through a hybrid (neurosymbolic) AI that leverages the strengths of both symbolic and subsymbolic models to overcome their respective limitations and achieve a more comprehensive understanding of the world.

In NLP research, this can be implemented in several ways, e.g., by injecting external knowledge into a deep neural neural network [14] in the form of embeddings (Fig. 4). Another recent neurosymbolic approach consists in a three-step normalization process [15] that first leverages linguistics to replace expressions like `shopping_for`, `bought`, and `purchasing` with their lemmas, e.g., `shop_for`, `buy`, and `purchase`, respectively. Next, deep learning models are used to cluster the resulting lemmas into primitives, e.g., `BUY`(x) is the cluster representing `shop_for`, `buy`, and `purchase`

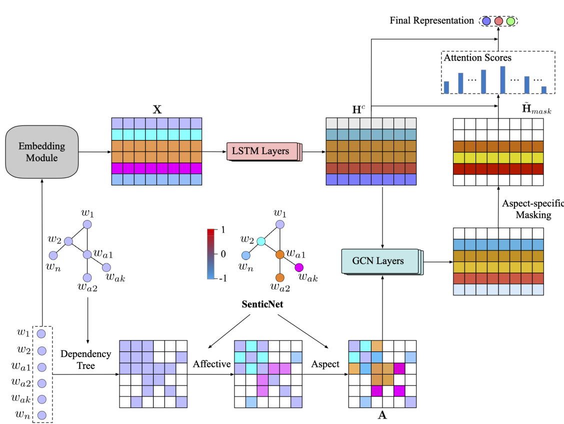

Figure 4. An example of neurosymbolic AI where (symbolic) commonsense knowledge is injected into a (subsymbolic) graph convolutional network [14].

and the likes of them (where x is the direct object acted upon by the primitive). Finally, logic is used to ground such primitives, e.g., `BUY` is defined in terms of `GET` and `GIVE`, which in turn are specified in terms of other superprimitives defining transfer of ownership.

In robotics, some researchers have emphasized the importance of physical embodiment and sensorimotor experiences in the development of intelligent systems. Such approaches, which go under the name of Embodied AI, are promising but they are still limited to very basic sensory experiences such as object manipulation. A similar approach can be taken in virtual worlds or in the metaverse, where AI could learn social commonsense, based on how people interact, and some sort of physical or spatial commonsense, such as gravity or the fact that you cannot go through walls. Additionally, an AI system could generate real-time virtual simulations to better perform causal reasoning and narrative understanding by grouding words into virtual objects and actions on the fly.

V. Similarity Measure

Because we have no better way of performing grounding, in computer science we use embeddings to represent data, e.g., text, audio, images, and videos, as vectors or data points in a multidimensional space. This mapping is learned from large amounts of data during a training process and it is usually focused on just one kind of similarity (usually semantic similarity based on word co-occurrence frequency). In order to enable *Parallel Analogy* (see Section III), we need to generate different representations for the same data based on different kinds of similarities.

Another problem is that we are still using very basic metrics to quantify the similarity between pairs of embeddings and, hence, perform classification. All such similarity metrics, e.g., Jaccard coefficient, Euclidean distance, and cosine similarity, blindly calculate distances in a multidimensional vector space without considering its topology. In the future, we need to adopt more topology-aware methods for calculating similarity in multi-dimensional vector spaces, e.g., Mahalanobis, Minkowski or Wasserstein distances and principal path methods [16].

These kernel methods are designed to discover smooth paths between objects in space by traversing a series of waypoints. One of their standout features is their ability to seek paths that pass through high-probability regions of the space, effectively navigating through geodesics influenced by the probability distribution of the sampled data. These methods also emulate the cognitive process where thinking involves transitioning from one concept to another while traversing regions of space with a high likelihood of encountering related concepts.

VI. Intention Awareness

Intention Awareness plays a crucial role in communication, as it enables individuals to anticipate and interpret the actions and behaviors of others, leading to more effective and empathetic interactions. Current AI models provide one-fits-all solutions without taking into account user beliefs, goals and preferences. Theory of mind should always be applied to better understand user's actions and queries. When this is not possible, user profiling in the form of persona or personality detection should be employed to generate more relevant actions or answers [17].

For the same reason, AI should also have enough commonsense knowledge, including a model of fundamental human beliefs, desires, and intentions, in order to minimize miscommunication and avoid unintended consequences (e.g., apocalyptic scenarios like accidentally wiping out humanity in the attempt to solve climate change). In other words, future AI systems should always try to understand what users are doing and why they are doing it. For instance, recent hybrid frameworks have tried to improve human-robot interaction by modeling *Intention Awareness* in terms of motivational and affective processes based on conceptual dependency theory [18].

Finally, recent attempts to augment the human decision-making process, especially in dynamic and time-sensitive scenarios such as military command and control, game theory, home automation, and swarm robotics, have focused primarily on environmental details such as positions, orientations, and other characteristics of objects and actors of an operating environment (situation awareness). However, a significant factor in such environments is the intentions of the actors involved [19]. While creating systems that can shoulder a greater portion of this decision-making burden is a computationally intensive task, performance advances in modern computer hardware bring us closer to this goal.

VII. Trustworthiness

Last but not least, *Trustworthiness* is a key pillar that measures the degree to which AI systems, models, and algorithms can be relied upon to perform as intended, make accurate and ethical decisions, and avoid harmful consequences. It is a concept closely related to *Intention Awareness* (see Section VI), but also explainability and interpretability. Explainability allows an AI model to generate descriptions of its decision-making processes in order to enable the user to make informed modifications to the outputs or even to the model itself in a human-in-the-loop fashion. Interpretability, in turn, enables users to understand the inner workings of an AI model, e.g., by identifying which input features have the most impact on its output or by assessing how changes in input variables affect the model's predictions or by leveraging a confidence score to gauge how confident the AI model is about its own output.

According to one theory of trust, trust is grounded in probabilities that a trustor A attributes to his/her own beliefs about the behavior and competences of a trustee B with respect to the performance of some action ϕ relevant to a goal G. Where n denotes the probability threshold value and m denotes the probability value that A attributes to his/her trust-relevant beliefs, there will be a trust relation between A and B if and only if $m \geq n$. It is at least arguable that this probability threshold value will be met, given the twin phenomena of hallucination and behavioral drift. In any case, we believe that trust is more than a matter of satisfying probability threshold conditions (i.e., $m \geq n$). We can define *Trustworthiness* as a 5-ary relation R(A, B, ϕ, ψ, G), consisting of five relata: the trustor A, a trustee B, some action ϕ to be performed, some G-relevant attribute ψ that may be judged by A as absent or present in B during B's performance of ϕ, and a goal G that makes the performance of ϕ desirable [20]. Indeed, trust is a mental state that A holds toward B with respect to the performance of some G-relevant ϕ.

If the goal G is NLP, actions ϕ_i are typical NLP tasks such as sentiment analysis and dialogue generation, and G-relevant attributes ψ_i are qualities such as explainability and interpretability. All other things being equal, if B_1 possesses each G-relevant attribute ψ_i in greater abundance than B_2, we have *pro tanto* reasons to have greater trust in the former than the latter.

Intuitively, even if real parrots or stochastic ones (LLMs) produce appropriate linguistic responses to task-related ϕ_i prompts, we would not deem their linguistic behavior trustworthy unless they possess the relevant natural language understanding. Meaning involves a relation between the linguistic form of data and an extralinguistic reality that is distinct from language. Where M denotes meaning, E denotes the form of natural language expressions, and I denotes communicative intent, this relation may be formally represented as $M \subseteq E \times I$ [2]. M contains ordered pairs (e, i) of natural language expressions (e) and communicative intents (i). Understanding may be interpreted as the process of retrieving i, given e. Since LLMs are pretrained on large datasets and meaning cannot be learnt from linguistic form (e) alone, however impressive their transformer architecture might be, LLMs will necessarily lack the relevant intentionality. Such a limit can result in hallucinatory responses from LLMs, if e and i are not directly associated in the pertaining datasets. Hence, humans have to watch over them and correct them in mission-critical tasks.

Ensuring *Trustworthiness* requires collaboration among AI experts, ethicists, and policymakers. It involves a combination of technical measures, ethical considerations, and transparency initiatives. As AI continues to play an increasing role in various aspects of society, building and maintaining trust in AI technologies is essential for their responsible and sustainable integration.

Conclusion

The pursuit of automating tedious or repetitive tasks has a rich history, with origins tracing back to Ancient Egypt and the Greek Empire. Among the earliest documented works on automation is the “Book of Ingenious Devices”, published in 850 by the Banu Musa brothers. While we have made significant strides since those times, thanks to advancements in mathematical modeling, we now face the challenge that mere mathematics alone may not suffice to model the intricate processes by which the human brain encodes and decodes meaning for complex tasks, including intuitive decision-making, sense disambiguation, and narrative comprehension.

In this work, we proposed a novel approach to AI that centers on humanity, characterized by seven essential features or pillars. In the future, we plan to define best practices for abiding by such pillars. For example, current post-hoc interpretability methods may not be the best way to implement *Trustworthiness* as they simply find correlations between inputs and outputs of an AI model without really explaining its inner workings. Similarly, there is no point in having a confidence score if this is calculated based on the wrong parameters. Moreover, we need to define how to evaluate explainability in terms of qualities such as plausibility, i.e., the extent to which an explanation resonates with and is deemed acceptable by a human audience, and faithfulness, i.e., the extent to which the explanation accurately reflects the model’s decision-making process.

As we look ahead, it is imperative to foster the development of human-in-the-loop and human-in-command systems, integrating human participation in AI through paradigms such as active learning and decision intelligence. We need to develop clear guidelines and principles for AI development which prioritize human values, fairness, accountability, transparency, and privacy, and which should be integrated into the design process from the outset. We need to conduct regular audits of AI algorithms to detect and mitigate biases, errors, and ethical concerns.

Finally, we need to implement and enforce regulations and governance mechanisms that define the boundaries of AI usage, protect individual rights, and foster moral AI practices. By implementing these strategies, society can work toward ensuring that AI technologies are developed and used in ways that empower individuals and align with ethical values. Balancing technological progress with human agency and values is essential for the responsible advancement of AI. If we do not engineer it well, in fact, AI could very much end up being like plastic: a great invention that made our life easier about a century ago, but which is now threatening our own existence.

REFERENCES

1. R. Bommasani et al., "On the Opportunities & Risks of Foundation Models," *arXiv preprint arXiv:2108.07258*, 2021.
2. E. M. Bender and A. Koller, "Climbing towards NLU: On Meaning, Form, & Understanding in the Age of Data," *ACL*, 2020, pp. 5185–5198.
3. M. Amin et al., "A Wide Evaluation of ChatGPT on Affective Computing Tasks," *arXiv preprint arXiv:2308.13911*, 2023.
4. R. Mao et al., "GPTEval: A Survey on Assessments of ChatGPT and GPT-4," *arXiv preprint arXiv:2308.12488*, 2023.
5. Y. Shen et al., "ChatGPT & other large language models are double-edged swords," *Radiology*, vol. 307, no. 2, 2023, p. e230163.
6. L. Chen, M. Zaharia, and J. Zou, "How is ChatGPT's behavior changing over time?" *arXiv preprint arXiv:2307.09009*, 2023.
7. V. Sorensen et al., "Mood of the Planet: Challenging Visions of Big Data in the Arts," *Cognitive Computation*, vol. 14, no. 1, 2022, pp. 310–321.
8. L. Floridi et al., "AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations." *Minds Mach*, vol. 28, 2018, pp. 689–707.
9. E. Cambria et al., "Sentiment Analysis is a Big Suitcase," *IEEE Intelligent Systems*, vol. 32, no. 6, 2017, pp. 74–80.
10. X. Zhang, R. Mao, and E. Cambria, "A Survey on Syntactic Processing Techniques," *Artificial Intelligence Review*, vol. 56, 2023, pp. 5645–5728.
11. R. Mao et al., "A Survey on Semantic Processing Techniques," *Information Fusion*, vol. 101, 101988, 2023.
12. M. Minsky, *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*, Simon & Schuster, New York, 2006.
13. E. Cambria, D. Olsher, and K. Kwok, "Sentic Activation: A Two-Level Affective Common Sense Reasoning Framework," *AAAI*, 2012, pp. 186–192.
14. B. Liang et al., "Aspect-Based Sentiment Analysis via Affective Knowledge Enhanced Graph Convolutional Networks," *Knowledge-Based Systems*, vol. 235, no. 107643, 2022.
15. E. Cambria et al., "SenticNet 7: A Commonsense-based Neurosymbolic AI Framework for Explainable Sentiment Analysis," *LREC*, 2022, pp. 3829–3839.
16. M. Ferrarotti, W. Rocchia, and S. Decherchi, "Finding Principal Paths in Data Space," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 8, 2019, pp. 2449–2462.
17. L. Zhu et al., "PAED: Zero-Shot Persona Attribute Extraction in Dialogues," *ACL*, 2023, pp. 9771–9787.
18. S.-B. Ho et al., "Knowledge Representation for Conceptual, Motivational, and Affective Processes in Natural Language Communication," *BICS*, 2023.
19. N. Howard and E. Cambria, "Intention awareness: Improving upon situation awareness in human-centric environments," *Human-centric Computing and Information Sciences*, vol. 3, no. 9, 2013.
20. M. Chen, "Trust & trust-engineering in artificial intelligence research: Theory & praxis," *Philosophy & Technology*, vol. 34, no. 4, 2021, pp. 1429–1447.

Erik Cambria is a Professor at NTU, where he also holds the appointment of Provost Chair in Computer Science and Engineering. His research focuses on neurosymbolic AI for explainable sentiment analysis in domains like social media monitoring, financial forecasting, and AI for social good. Contact him at cambria@ntu.edu.sg.

Rui Mao is a Research Fellow at Nanyang Technological University, studying conversational AI. His research interests include computational metaphor processing, affective computing and cognitive computing. Contact him at rui.mao@ntu.edu.sg.

Melvin Chen is a Faculty Member of both the Philosophy programme and the University Scholars Programme and the Assistant Chair (Communications & Outreach) at the School of Humanities at NTU. Melvin's research interests include the philosophy of AI, creative cognition research, aesthetics, ethics and metaethics, philosophy and literature. Contact him at melvinchen@ntu.edu.sg.

Zhaoxia Wang is an Associate Professor of Computer Science (Practice) in the School of Computing and Information Systems of Singapore Management University. Her research interests include natural language processing, machine learning, sentiment analysis, causal reasoning, and intelligent robots. Contact her at zxwang@smu.edu.sg.

Seng-Beng Ho is a Senior Scientist at the Institute of High Performance Computing, A*STAR, Singapore. He has published many AI-related papers in international journals and conferences in which he presents principled and fundamental theoretical frameworks that are critical for building truly general AI systems. Contact him at hosb@ihpc.a-star.edu.sg.