

The impact of artificial intelligence (AI) on translation: An overview

Stefania M. MACI, University of Bergamo, Italy

To be operational, Artificial intelligence (AI) uses large language models (LLMs) in the domain of Natural Language Processing (NLP). A significant advancement in this field is Transformer architecture, a deep learning model initially developed for machine translation. The advent of AI conversational interfaces, such as chatbots, has triggered scientific, philosophical, political, social, ethical and legal debates about the pros and cons of AI for humanity, also focusing on the 'ability' of AI to accurately translate a wide range of (specialised) texts. These chatbots, powered by LLMs trained on extensive data sets from books and the Internet, are designed to generate coherent and contextually relevant text by understanding and predicting language patterns. Hence, after a diachronic overview of AI and automatic translation, the aim of this paper is to show how chatbots "translate" specialized texts, comparing different chatbots to identify their similarities and differences. While Neural Machine Learning (NML) shows promise, it still struggles to produce consistently satisfactory results.

Keywords: AI; Artificial Intelligence; Artificial Neural Networks; Automatic Translation

1. Introduction

The classical approach to Artificial Intelligence (AI), also known as Good Old-Fashioned AI (GOFAI), traditionally worked on knowledge in very structured ways, using formal logic and symbolic representations and explicit models of the world (such as *Wikipedia*). Contrary to this, nowadays, generative AI learns to generate appropriate outputs based on huge amounts of training data that are 'prepared' but not structured. The learning models of a deep neural network develop complex representations of data, difficult to express in symbolic or logical terms and often considered 'black boxes.' From this perspective, such systems can be 'oracular' because the process that leads to the production of the output is never explicable according to a traditional architectural paradigm (Roncaglia, 2023).

1.1. From strong AI to generative neural networks

The rise of AI coincides with Turing's seminal work "Computing machinery and intelligence" published in 1950 (Turing, 1950). By developing the *imitation*

game test,² Turing tries to determine whether a machine can exhibit intelligent behaviour equivalent to that of a human being. Although he knew there would be objections³ to the idea that machines could be intelligent, he nevertheless believed that human intelligence might be replicable through computational means. If so, machines do not need to be programmed to perform specific tasks: they can learn from experience and be trained to develop 'intelligence'.

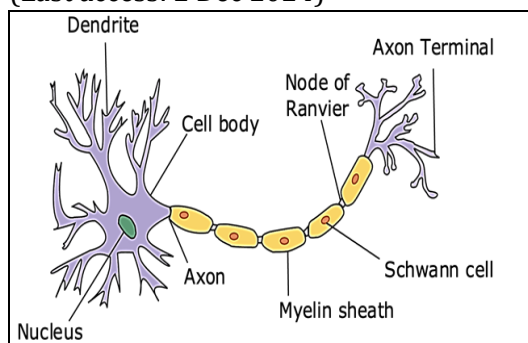
AI, or better 'strong' AI, was established as a formal field of study with the 1956 Dartmouth Summer Research Project. Here, the idea that machines could simulate human intelligence, which would be 'artificial intelligence' (Moor, 2006), was explored. This notion of AI lay in the ideas that (a) our intelligence is mainly linguistic, as it is mainly manifested using language; and (b) language is a system governed by rules that can be formulated so precisely and rigorously that they can be programmed and used with computer systems. These ideas, developed from the 'linguistic turn' in formal logic philosophy, centred around language (Roncaglia, 2023), were further supported by Chomsky's generative transformational grammar (Chomsky, 1969), which hypothesised the existence of a deep grammatical structure common to different languages and analysable in rigorously formal terms. Since the 1970s, however, these premises have been called into question (Roncaglia, 2023). From a strictly technological point of view, as early as the mid-1970s, many researchers began to abandon the idea of strong AI in favour of a much broader but less ambitious range of research directions, grouped under the umbrella term 'weak AI'.⁴

Amongst the various studies put forward in the 1956 Dartmouth Summer Research project, one about neural networks explored the possibility of replicating certain aspects of the functioning of our brains through networks of artificial neurons.⁵

This idea was originated in the 1940s by McCulloch and Pitts (McCulloch & Pitts, 1943), who proposed to look at the neuron as a sort of computation machine: on the basis of the information collected by the dendrites and processed by the nucleus, the neuron produces an output transmitted to other neurons by the axon and via the synapses. Similarly, McCulloch and Pitts' model, represented as a binary threshold device (McCulloch & Pitts, 1943), maintained that each neuron receives multiple

Figure 1.

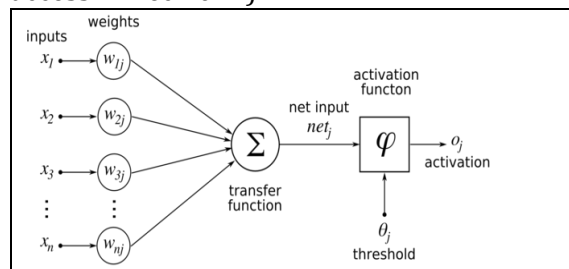
Schematic Representation of a Neuron (available at: commons.wikimedia.org) (Last access: 2 Dec 2024)



binary inputs (either 0 or 1), which are then summed. This model, also known as the M-P neuron (after the initials of their names), uses simple logical operations like AND, OR and NOT to mimic neural activity and demonstrates how neural networks can perform any logical function or computation. In this way neurons seem to behave like logical operators (e.g., an AND neuron will return a value of 1 if all the inputs received have a value of 1, while an OR neuron will return a value of 1 if at least one of the inputs has a value of 1, and so on) (Chakraverty et al., 2019). McCulloch and Pitts' model is certainly an extreme simplification of a much more complex reality, but it was crucial in establishing the idea that neural networks could be used to represent and compute complex functions. It represents the first and fundamental step in computational neuroscience and artificial neural networks (Marsalli, 2007).

A further implementation of Figure 2.

McCulloch and Pitts' model is constituted by the so-called *perceptron*, a model developed by Rosenblatt (1958). Based on this model, the perceptron accepts inputs with values other than 0 or 1. To these inputs different *weights* are assigned, in accordance with the importance they have in contributing significantly to the final output (IBM, 2024). Different from the M-P neuron model (where the sum of the inputs is predictable and logic-deterministic), in Rosenblatt's model the sum of the *weights* activates the perceptron on a *probabilistic* threshold, which can be modified either by the programmer or according to the input received from other perceptrons. The possibility of modifying the threshold value and the *weights* of the various inputs during training allows the perceptron to 'learn'. It is based on this probabilistic model that, nowadays, artificial neural networks are operated. This allows finer training of the neural network, reducing determinism (Roncaglia, 2023).



Perceptron-based networks and neural networks typically comprise an input layer, which receives information from the outside world, an intermediate layer, which processes and 'learns', and an output layer, which returns the result of the network's work to the outside world. Usually, a neural network is trained on the basis of a comparison between produced and desired output, progressively minimising the level of deviation (or error): a *backpropagation* errors algorithm allows dynamically modifying the weights within the network according to the distance between the result obtained and that expected, achieving learning based on reinforcement feedback (when the distance from

the expected result decreases) or weakening (when it increases). This feedback can be human and/or automatic, and it can also be entrusted to other neural networks. This model is also called *deep neural networks* or *deep learning systems* because, during training, the network may change weights and values autonomously, without the person programming it knowing the states of the individual neurons comprising it. The ability to build this type of model makes systems particularly effective in tackling complex problems such as image recognition, natural language understanding and speech recognition (IBM, 2024; Roncaglia, 2023).

1.2. Generative Neural Networks

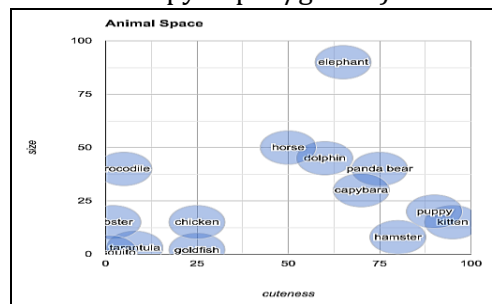
Generative neural networks such as GPT generate texts in response to user prompts. These systems are based on NLP models and start with the selection and pre-processing of a large text corpus, which is divided into tokens (language units). During training, the neural network learns to predict the next token based on previous ones. This is done thanks to the initial corpus and the gradual adjustment of the values and weights assigned to tokens, resulting in the Large Language Model (LLM), a

statistical-probabilistic model correlating tokens, each represented by a large matrix of numerical values. In this way, each token is associated with an abstract and multidimensional space that expresses in a numerical way (the *vectorisation* process; Cf. Kaushik, 2024) contexts of use and the relationships of tokens in the corpus: tokens with similar ‘uses’ and similar meanings correspond to vectors that have numerical values quite close to each other. For instance, if we want to find what association exists between animal ‘cuteness’ and animal size, ‘size’, the vectorisation of these qualities allows us to visualise it (Parrish, 2024) as can be seen in Figure 3.

The construction of vectors for each token provides the *embedding*, i.e., the representation of usage patterns in the language as numerical values (Kennedy, 2024). The model thus constructed is then used to generate a process of response generation by AI, thanks to a mechanism called *sequence-to-sequence*, in which a sequence of input symbols is transformed into a sequence of output symbols.

Figure 3.

Two-Dimensional Visualisation of ‘Cuteness’ and ‘Size’ in Animal via Vectorization (Retrieved 2nd December 2024 from: ipynb.pub/github)



Originally, recurrent neural networks (RNNs) were used for similar tasks, processing information word by word. However, RNNs had problems with 'semantic memory', i.e., difficulties in maintaining the coherence of meaning in long texts. A significant advance was made in 2017 with the introduction of transformers (Devlin et al., 2019; Turner, 2024; Vaswani et al., 2017), a network architecture that is more efficient than RNNs, as it allows the context of the tokens to be taken into account through an attention mechanism, by weighing the values of the vectors of each token against the values of every other token in context. This method improves semantic consistency and the handling of long inputs by re-iterating the *attention mechanism* with a *multi-headed attention* technique, which repeatedly takes into account the weights the inputs have syntactically and semantically: this facilitates the disambiguation of polysemous words, works much better with long inputs and produces systems with a better 'semantic memory' (Vaswani et al., 2017).

Transformer-based architecture is applied several times in succession, in both the encoding of the input and the production of the response. Two different modules, called 'encoders' and 'decoders', respectively, may be involved in this process. For instance, in most generative artificial intelligences dedicated to translation, the encoder deals specifically with the representation of text received as input by vectors, and the decoder with the generation of output, starting from the representation generated by the encoder (Roncaglia, 2023). However, transformers that focus mainly on the aspect of text representation and analysis (and therefore only use encoders) or transformers that focus mainly on text generation (and only use decoders) are also possible (Roncaglia, 2023). The choice of the most functional model (encoder-decoder, encoder only, or decoder only) depends partly on the researcher's objectives. For example, BERT (Bidirectional Encoder Representations from Transformers), one of the first models proposed in 2018 (Devlin et al., 2019), aimed to create an LLM capable of 'predicting' both a token in its context and the next sentence in a given context (NSP—Next Sentence Prediction).

It is, therefore, a pure encoder model that is effective in text analysis. This type of architecture works well in sentiment analysis or in the identification of different 'agents' in a dialogue, or even in the creation of summaries or paraphrases of a text or its semantic analysis. On the other hand, GPT models, created by OpenAI, mainly work on systems that consist only of decoders and are primarily focused on text production. Their goal is to predict the next token from previous ones. The words produced by the decoders themselves become part of the input used to produce the next word. This makes it possible to generate responses that maintain syntactic and semantic coherence, even if they are much longer than the initial prompt. Interestingly, the GPT model does not always select the 'most relevant' word, but rather, through a stochastic

component, also words with slightly lower scores to make the text more varied or interesting (Steel, 2023).⁶

After the unsupervised learning phase, the model can be refined using supervised learning and reinforcement learning techniques. In supervised learning, the model is trained with input-output pairs that have been validated by humans. In reinforcement learning, the system's responses are evaluated and then 'rewarded' or 'punished' to improve performance. User interactions, such as feedback, can be used for further training. The system can also use other AI to analyse the output (e.g., Generative Adverse Networks—GANs) and apply filters to avoid unacceptable responses, thus reducing errors and inappropriate behaviours, such as the use of offensive, violent or sexually explicit language or hallucinations (Roncaglia, 2023).

After the unsupervised learning phase, the model can be refined using supervised learning and reinforcement learning techniques. In supervised learning, the model is trained with input-output pairs that have been validated by humans. In reinforcement learning, the system's responses are evaluated and then 'rewarded' or 'punished' to improve performance. User interactions, such as feedback, can be used for further training. The system can also use other AI to analyse the output (e.g., Generative Adverse Networks—GANs) and apply filters to avoid unacceptable responses, thus reducing errors and inappropriate behaviours, such as the use of offensive, violent or sexually explicit language or hallucinations (Roncaglia, 2023).

2. AI and translation

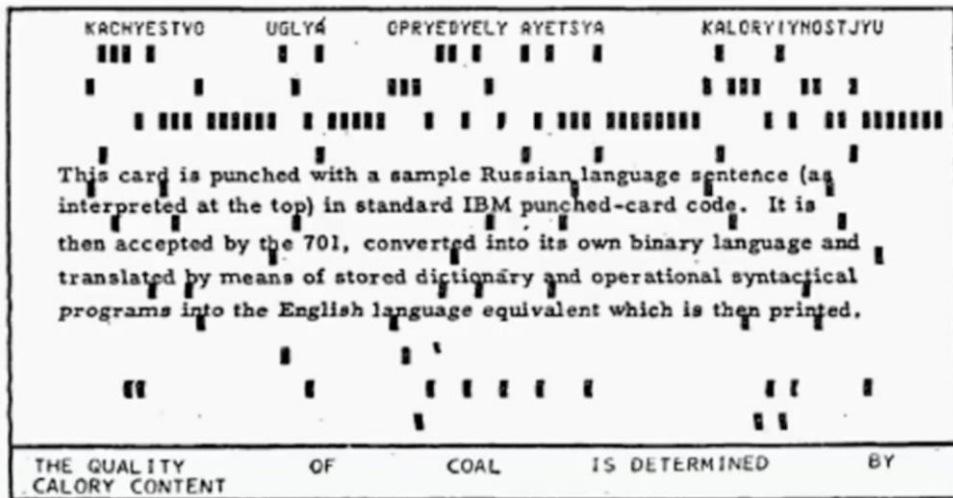
Automatic translation (AT) is one of the tasks of NLP and its history retraces the history of AI as overviewed above.

The rise of AT dates back to the 1950s when, in the middle of the Cold War, the URSS and the US were competing for the invention of a machine for mechanical translation. In 1956, Georgetown University and IBM carried out the so-called Georgetown-IBM experiment, which contributed to the success of machine translation characterised by rule-based models (Jones, 1994; Kuka & Se, 2023). Thanks to the Georgetown-IBM machine (see Figure. 4 below), 43 sentences were translated from Russian into English in a public demonstration of mechanical translation using punched cards. The demonstration was a big success and it laid the foundations and set the main direction of Machine Translation (MT) research for the following years (Kuka & Se, 2023). Yet, in 1966, the Automatic Language Processing Advisory Committee (ALPAC) published a report which strongly criticised AT, because it was believed that, given the technology of that time, it was impossible to have high translation performance with automatic tools. For this reason, their recommendation was

that research on such development should not be funded further (ALPAC, 1966; Kay, 1973). This halted research on AT until the 1980s.⁷

Figure 4.

The Punched Card of the Georgetown-IBM Experiment (Retrieved 2 December, 2024 from <https://www.turingpost.com/p/llmshistory1>)



From the 1980s onwards, MT experienced a notable resurgence, with European and Japanese contributions playing key roles. The European Commission adopted practical production systems and, with the Eurotra project, focused on multilingual translation using a standardized transfer framework. In Japan, multiple teams were actively involved, with some MT products already available in the market (Jones, 1994). The MT efforts of the time were based on the idea that highly accurate translations could be achieved, particularly for specialized tasks, with or without the involvement of editors or users, thanks to statistical models based on big bilingual data to learn translation patterns which could *predict* correct translation. These systems reflected the advancements in NLP, especially in grammar selection and modular system design (Jones, 1994).

The late 1980s saw a significant increase in research focused on the lexicon, driven by the key role it plays in a grammatico-logical approach, the demands of multilingual MT and the challenges of transportability, customization and knowledge acquisition for specific applications (Jones, 1994). Commercial dictionaries in machine-readable formats began to be utilized, which in turn led to the use of text corpora to validate, enhance or customize initial lexical data. In these years, translation memories, i.e., databases that store previously translated sentences and their corresponding translations, began to be used to

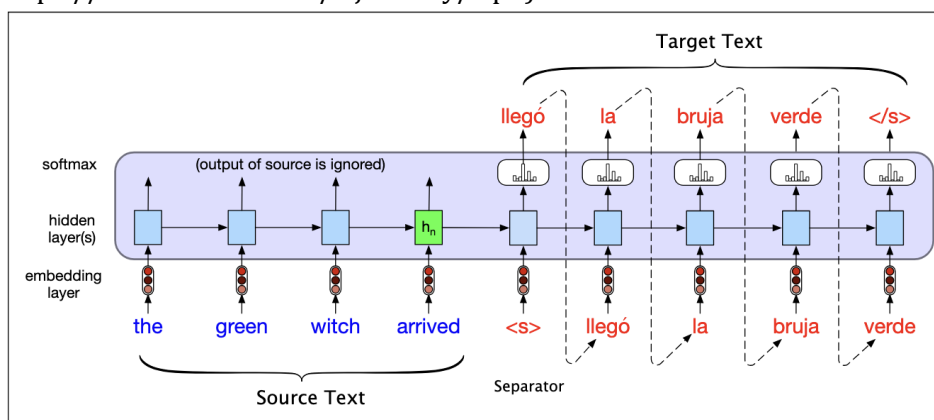
speed up the translation process and ensure consistency in terminology and style. These databases were used alongside the rise of computer-aided translation (CAT) tools and parallel corpora, which became more popular during the 2000s. While translation memories were used for practical translation tasks, parallel corpora also began to be used for research and development purposes (Jones, 1994).

The revolution in AT began in the 2010s with Neural Machine Translation (NMT), which uses neural networks to learn complex patterns in language. Google translate, released as a statistical MT in 2016, transited to a neural MT premise (Sommerland, 2021) and has since then significantly improved the quality of MT by using Transformer architecture.

The introduction of Transformer architecture marked a significant turning point in the field of MT. Introduced in 2017 (Vaswani et al., 2017), this innovative model has quickly become the standard for many MT systems. At the heart of Transformer architecture is the encoder-decoder paradigm. The encoder carefully processes the input sequence and converts it into a series of vectorial representations, which then serve as the basis for the decoder, iteratively generating an output sequence, token by token, using both the encoder's output and its own previously generated tokens. In other words, the Transformer tokenizes and embeds the data on the neural plane of a language, transferring the prompt, token by token, to the other language (see Figure. 5 below).

Figure 5.

Translating a Single Sentence with the Basic RNN Version of Encoder-Decoder Approach to Machine Translation (Jurafsky & Martin, 2024, p. 205. Available at: <https://web.stanford.edu/~jurafsky/slp3>)



The Transformer's *attention mechanism*, which assigns specific *weights* to the tokens based on the context created so far with the decoding, plays a

fundamental role in predicting and translating the next token. This capability is fundamental for capturing wide-ranging dependencies within input and output sequences, a crucial aspect of effective translation (Jurafsky & Martin, 2000).

3. Translating texts with AI

As seen above, NLP and Transformer architecture brought about a revolution in AT to such an extent that “the translation quality of NMT has significantly improved and exceeded that of SMT [statistical Machine Translation] (Zhao et al., 2023). However, even though AT tasks are based on LLMs and Transformers, the software works on segments and has no ability to see the whole. The resulting text is readable but there are hidden mistakes, which are more evident in literary translations, where the author’s stylistic creativity cannot be conveyed by NMT. This requires a more conceptual post-editing process than a grammar-based one, done by human translators who have the ability to perceive nuances that these systems do not yet have (Brusasco, 2018). Indeed, while AT is predictive, human translation is interpretative. However, it is believed that in the case of specialised translations, information technology developers can, optimistically, develop predictive tools of high AT quality (Brusasco, 2018).

As claimed by Monteleone (2022), there exist a series of popular on-line translation portals, tools and chatboxes whose translation reliability is evaluated subjectively, and on the basis of language competence (cf. also Raffone, 2023). Despite the advent of Transformers, mistakes are still found. The following section shows what mistakes chatboxes can make when translating specialised texts.

3.1. Chatboxes translating (specialised) discourse

We asked three different chatboxes—Microsoft Copilot, Gemini, and ChatGPT—to translate first a tourism and then medical texts.

We first asked Microsoft Copilot, Google Gemini, and ChatGPT, to translate, from Italian into English, the first seven lines of the webpage *2 days in Rome: the itinerary* taken from the Italian Ministry of Tourism official website (<https://www.italia.it/it/lazio/roma/cosa-fare/weekend-roma>). In all three chatboxes, the translation from Italian to English was ok, but with some changes in the style. However, Microsoft Copilot added emojis (“Buon viaggio! 🌟”) and a promotional message (“For a detailed itinerary, consider visiting the *Sarà Perché Viaggio* website, which provides day-by-day recommendations and a walking map for your Roman adventure. Enjoy your weekend in the Eternal City! 🌟”), both of which were completely absent in the original web-text. What is more, Google Gemini, instead of simply translating the text,

offered us a full 2-day itinerary in Rome to which tips were added (e.g., “Tips: Purchase skip-the-line tickets to avoid lengthy queues and maximize your time”), which were not requested. Although these translations were ok, the fact that extra and promotional texts are added is questionable (not to mention the marketing operation!).

We then asked the same chatboxes to translate an excerpt taken from a medical textbook written by Cael (2010). The text to be translated from English to Italian was: “Directional terms are useful for describing the location of injuries as in ‘the client is experiencing soreness about 2 inches proximal to the left patella’” (Cael, 2010, p. 3). Its translation can be seen in Table 1 below (emphasis added):

Table 1
Medical Text Translations in Different Chatboxes

Microsoft Copilot	Google Gemini	ChatGPT
I termini direzionali sono utili per descrivere la posizione delle lesioni, come in “ il cliente sta sperimentando un dolore a circa due pollici prossimale alla rotula sinistra.”	I termini direzionali sono utili per descrivere la posizione di lesioni, come in “ il paziente lamenta dolore circa due pollici prossimalmente alla rotula sinistra.”	I termini direzionali sono utili per descrivere la posizione delle lesioni, come ad esempio, “ il cliente sta avvertendo dolore circa due centimetri prossimali alla rotula sinistra.”

Clearly, although the text is syntactically correct, there are cultural mistakes. First, the idea that a patient is a ‘client’ is US culturally based. Only Google Gemini translates it correctly, carrying out cultural mediation. Secondly, there is an obvious cultural gap in translating ‘inch’. As we know, the metric system is completely different from the imperial one. If Microsoft Copilot and Google Gemini correctly translated ‘inches’ with ‘pollici’, the translated text is culturally wrong, because this should have been rendered in ‘centimetri’. ChatGPT correctly mediates ‘inches’ in ‘centimetri’, but the measurement is wrong: 1 inch corresponds to 2.54 centimetres. So, the correct translation would have been ‘5.08 centimetres’. Let’s now suppose that Italian students in Medicine and Surgery need to study this textbook but do not have high competence in English and so ask any of these chatboxes to translate the text into Italian. In this case, the mistake cannot be considered trivial: the translation from ‘inches’ to ‘pollici’ does not help students who may have no knowledge of the imperial measurement system; in the case of the translation from ‘inches’ to ‘centimetres’ without adjusting the corresponding measure, students may be induced to make lethal medical mistakes.

4. Concluding Remarks

In this paper, after an overview of the rise of AI and its use in MT, we highlighted that NMT is characterized by subtle errors (which can be harder to detect than errors made by SMT). This is especially true for users without linguistic training and lack digital and quality awareness: they may mistakenly trust the output as error-free. Yet, as we have seen above, there exist errors with NMT, such as wrong or misleading meanings, omissions, extra text and hallucinations, to such an extent that the quality of translations can be unreliable. Although the technology continues to evolve, it is evident that NMT is the first tool capable of learning from language experts. This means that ownership of NMT must shift from technologists to linguists and translators, who are better equipped to guide its development in real-world applications. This calls for appropriate strategies, not only such as AT, combined with human post-editing, but also mastering other Machine translation tools: new tools demand new competencies, deeply embedded deeply within cultural and professional contexts where precise communication is critical (Yamin et al, 2023). This is possible if we reassign to these tools the role they have.

Notes

1. In the *imitation game*, an interviewer interacts with two entities (a human being and a machine) via a text-interface and must identify the human and the machine (Turing, 1950).
2. These were: (a) the *theological objection*: thinking is an immortal soul's function, which machines cannot possess; (b) the *heads in the sand objection*: if machines could think, the output would be unsettling; (c) the *arguments from consciousness objection*: machines cannot have conscious experiences and therefore cannot truly 'think'; (d) the *mathematical objection*: a machine has limitations in what it can compute due to Gödel's incompleteness theorems.
3. Examples of 'weak' AI are Siri or Alexa (Chandler & Munday, 2020).
4. Neural networks represented a hybridisation between artificial intelligence and the reflections on the relationship between man and machine proposed by another of the research directions of the time, cybernetics (Roncaglia, 2023).
5. This process is controlled by a parameter called 'temperature'. When the temperature is high (e.g., 0.8), the system tends to be more creative and bring more variety to the answers. When the temperature is low (0), the system always selects the most predictable word, resulting in more accurate but less original answers (Steel, 2023).
6. At that time translations adopted rule-based systems characterized by dictionary-based word-for-word processing and operated with a set of

linguistic rules predetermined by linguists and experts—which meant that translating what is implicit could not be managed. However, in the 1960's the AT aims were not that of proposing a qualitative perfect output but rather that of minimising the error percentage in translation considering accuracy and fluidity (Jones, 1994).

The Author

Stefania M. Maci (Email: stefania.maci@unibg.it) (PhD, Lancaster University, UK) is Full Professor of English Language, Translation and Linguistics at the University of Bergamo, where she is Coordinator of the MA in Digital Humanities, Director of the Research Centre on Specialised Languages, and Scientific Coordinator of the Research Group on Digital Humanities. She served on the executive board of AIA (Associazione Italiana di Anglistica 2019-2023) and she is now serving on the executive board of CADAAD (Critical Approaches to Discourse Analysis across Disciplines). Her research is focused on the study of the English language in academic and professional contexts, with particular regard to the analysis of tourism and medical discourses.

References

- ALPAC, A. L. P. A. C. (1966). *Language and machines* (p. 138). National Academy of Sciences. <https://www.mt-archive.net/50/ALPAC-1966.pdf>
- Brusasco, P. (2018). La traduzione automatica—tradurre. *Tradurre. Pratiche, teorie, Strumenti*. <https://rivistatradurre.it/la-traduzione-automatica/>
- Cael, C. (2010). *Functional Anatomy: Musculoskeletal Anatomy, Kinesiology, and Palpation for Manual Therapists* (Wolters Kluwer).
- Chakraverty, S., Sahoo, D. M., & Mahato, N. R. (2019). McCulloch-Pitts Neural Network Model. In S. Chakraverty, D. M. Sahoo & N. R. Mahato (Eds.), *Concepts of soft computing: Fuzzy and ANN with programming* (pp. 167-173). Springer. https://doi.org/10.1007/978-981-13-7430-2_11
- Chandler, D., & Munday, R. (2020). *A dictionary of media and communication*. Oxford University Press.
- Chomsky, N. (1969). *Aspects of the theory of syntax*. MIT Press.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of deep bidirectional transformers for language understanding* (No. arXiv).

- IBM. (2024). *Cos'è una rete neurale?* | IBM. (2024, May 16). <https://www.ibm.com/it-it/topics/neural-networks>
- Jones, K. S. (1994). Natural language processing: A historical review. In A. Zampolli, N. Calzolari & M. Palmer (Eds.), *Current issues in computational linguistics: In honour of Don Walker* (pp. 3-16). Springer. https://doi.org/10.1007/978-0-585-35958-8_1
- Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (1st ed.). Prentice Hall.
- Jurafsky, D., & Martin, J. H. (2024). *Speech and language processing*. University of Colorado.
- Kaushik, V. (2024, May 4). *Understanding vectorization: Applications, benefits, and future trends*. <https://medium.com>
- Kay, M. (1973). Automatic translation of natural languages. *Daedalus*, 102(3), 217-230. <https://www.jstor.org/stable/20024156>
- Kennedy, C. (2024, January 24). *Are embeddings more than just tokenized text?* [Community]. <https://community.openai.com>
- Kuka, V., & Se, K. (2023, August 16). *The era of mechanical translation and how it crashed* (History of LLMs #1). <https://www.turingpost.com>
- Marsalli, M. (2007). *McCulloch-Pitts Neurons*. <https://mind.ilstu.edu>
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4), 115-133. <https://doi.org/10.1007/BF02478259>
- Monteleone, M. (2022). Zellig S. Harris' transfer grammar and its application with NooJ. In M. González, S. S. Reyes, A. Rodrigo & M. Silberstein (Eds.), *Formalizing natural languages: applications to natural language processing and digital humanities* (pp. 65-75). Springer Nature. https://doi.org/10.1007/978-3-031-23317-3_6
- Moor, J. (2006). The dartmouth college artificial intelligence conference: The next fifty years. *AI Magazine*, 27(4). <https://doi.org/10.1609/aimag.v27i4.1911>

- Parrish, A. (2024). *Understanding word vectors*. <https://ipynb.pub>
- Raffone, A. (2023). Combining critical linguistics methods and novel pedagogies: Digital storytelling and discourse analysis for social change. *International Journal of Language Studies*, 17(2), 1-24.
- Roncaglia, G. (2023). *L'architetto e L'Oracolo*. Laterza. <https://www.laterza.it/scheda-libro>
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386-408. <https://doi.org/10.1037/h0042519>
- Sommerland, J. (2021, March 24). This is how Google Translate actually works. *The Independent*. <https://www.independent.co.uk/tech>
- Steele, C. (2023). *Understanding OpenAI's temperature parameter*. Colt Steele. <https://www.coltsteele.com/tips>
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind, New Series*, 59(236), 433-460. <http://www.jstor.org/stable/2251299>
- Turner, R. E. (2024). *An introduction to transformers* (No. arXiv:2304.10557). arXiv. <https://doi.org/10.48550/arXiv.2304.10557>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30. <https://papers.nips.cc>
- Yamin, M., Setiawan, S., & Anam, S. (2023). Enhancing critical thinking to foster students' analytical capacity in academic writing. *International Journal of Language Studies*, 17(1), 53-70. <https://doi.org/10.5281/zenodo.7513369>
- Zhao, Y., Zhang, J., & Zong, C. (2023). Transformer: A general framework from machine translation to others. *Machine Intelligence Research*, 20(4), 514-538. <https://doi.org/10.1007/s11633-022-1393-5>