National Science and Technology Forum

S.E.T. for socio-economic growth

Opinion Piece by Ms Jansie Niehaus

Brave new world of Artificial Intelligence

What do we really want from Artificial Intelligence (AI)? Is ChatGPT dangerous? What values really make us human? The arrival of the super conversational chatbot might be a sign that humanity should agree on what it really wants, and soon.

I am really quite open-minded… I don't mind speaking with a robot, unless it looks like a cute toy and is placed in a hotel foyer; or if it is standing on a street corner in South Africa... Otherwise, I prefer to use a robot that recognises my accent or takes written input; that knows what the answers might be to my questions; is able to refer me to a real human if it can't resolve my issue; and most importantly: if I don't have to fight it – whether to convince it to do what I want it to, or regarding the accuracy of its information. In short, a robotic system of any kind really has to be an improvement on a human being. What about Autocorrect?: I use a mixture of English and Afrikaans to communicate, which leads to perpetual little arguments with Autocorrect in every message I type. I'm sure that we South Africans with our twelve (yes, now there are twelve!) national languages, all experience this frustration. When will someone develop an Autocorrect that learns quickly for multilingual communication?

Now a bot has arrived that can write us informative articles, essays, school assignments and poetry – all in perfect English. Will I have an argument with it? I thought it might be interesting to find out more, how it has come about, what it is trying to do, and ask the question: SHOULD WE USE IT?

What is ChatGPT?

ChatGPT is a chatbot. AI chatbots are not new and it's not unusual to ask them questions. They are assistants on some websites and give just-in-time advice. It also seems that people are getting used to giving prompts that are more understandable to automated systems. While developers train AIs, AIs seem to be training users too.

Nor is learning perfect English a new development for AI. It has been correcting our British English in South Africa to American English for decades now! But seriously, if you get your settings right, spelling and grammar correcting software is extremely useful and we wouldn't want to live without such AI anymore, would we?

Since chatbots and virtual assistants have had to learn human language through natural language processing (NLP) to interact with humans, it was actually a small step to imitating articles, essays and poems written by humans. AI is good at reading, discerning patterns and classifying text according to characteristics – better than any of the school children and students whom we would like to teach creative or scientific writing! There are also enormous amounts of data for AI to use on the internet

to 'deep-learn' from, classify, and imitate. Nevertheless, the outcomes of AI's 'creative processes' do come as a shock, precisely because they look like genuine human products.

It seems logical now, that sooner or later, something like ChatGPT was bound to come along, and yet, it has taken the world by storm, especially the academic and education world. Many of us are concerned about its use by students and learners. Would they simply take advantage of ChatGPT to do their academic work for them, with the teachers and lecturers none the wiser? What are the implications of its availability to some and not to others?

As with other applications of Artificial Intelligence (AI), we can't wish it away now. Many a genie has escaped its bottle and we cannot put them back. We have to find ways to work with, and around, them. While ChatGPT has been shocking educators and academics, Google also launched its own sophisticated chatbot not so long ago, in 2021: LaMDA: our breakthrough conversation technology (blog.google) (more about that here below). There are other examples.

If you search for ChatGPT: The 'Search Engine Journal' comes up among the first items: ChatGPT: What Is It & How Can You Use It? (searchenginejournal.com). Here follow some useful extracts of the text on that web page:

"OpenAI introduced a long-form question-answering AI called ChatGPT that answers complex questions conversationally. It's a revolutionary technology because it's trained to learn what humans mean when they ask a question."

In short, ChatGPT is a phenomenal technological development. It is also trained to be a people-pleaser! For one thing, it '**listens'** to what you ask of it and improves itself so that it can give you answers that you'll like.

What do people really want?

It's also meant to give accurate information, but it turns out that these requirements – pleasing people and giving them the truth - contradict each other. The developers could have known this, because humans in the form of politicians also find those two things impossible to do at the same time! Thus there are hilarious examples of ChatGPT providing incorrect information in impeccable English. One of them is the answer to the question: What is the largest mammal that lays eggs? ChatGPT said: Elephants! And proceeded to explain that they are large eggs and other sentences seeming to follow nicely on the opening sentence. OpenAI had collected them before its launch and have warned about this risk:

"The experience of Stack Overflow moderators with wrong ChatGPT answers that look right is something that OpenAI, the makers of ChatGPT, are aware of and warned about in their announcement of the new technology."

The OpenAI announcement offered this caveat:
"ChatGPT sometimes writes plausible-sounding but incorrect or nonsensical answers.
Fixing this issue is challenging, as:
(1) during RL training, there's currently no source of truth;
(2) training the model to be more cautious causes it to decline questions that it can answer correctly; and
(3) supervised training misleads the model because the ideal answer depends on what the model knows, rather than what the human demonstrator knows."

'There's currently no source of truth.' There you have it! A true post-truth technology. I will definitely be arguing with it about the human existential dilemma and our relationship with the truth.

Large Language Models

"ChatGPT is a large language model (LLM) chatbot developed by OpenAI based on GPT-3.5. It has a remarkable ability to interact in conversational dialogue form and provide responses that can appear surprisingly human. Large Language Models (LLMs) are trained with massive amounts of data to accurately predict what word comes next in a sentence.

"Reinforcement Learning with Human Feedback (RLHF) is an additional layer of training that uses human feedback to help ChatGPT learn the ability to follow directions and generate responses that are satisfactory to humans. It was discovered that increasing the amount of data increased the ability of the language models to do more.

"According to Stanford University: "GPT-3 has 175 billion parameters and was trained on 570 gigabytes of text. For comparison, its predecessor, GPT-2, was over 100 times smaller at 1.5 billion parameters. This increase in scale drastically changes the behavior of the model — GPT-3 is able to perform tasks it was not explicitly trained on, like translating sentences from English to French, with few to no training examples.

"LLMs predict the next word in a series of words in a sentence and the next sentences – kind of like autocomplete, but at a mind-bending scale. This ability allows them to write paragraphs and entire pages of content. But LLMs are limited in that they don't always understand exactly what a human wants. And that's where ChatGPT improves on state of the art, with the aforementioned Reinforcement Learning with Human Feedback (RLHF) training."

Harmless words?

The developers of ChatGPT were aiming to make it both 'truthful and harmless', i.e. not 'toxic'. I am not sure what they meant by that, but one could perhaps assume: not racist, not sexist, or not offensive in other ways? It will avoid questions that use offensive words and avoid using them, but how well is it able to pick up offensive nuances and implications?

"A March 2022 research paper titled *Training Language Models to Follow Instructions with Human Feedback* explains why this is a breakthrough approach:

"This work is motivated by our aim to increase the positive impact of large language models by training them to do what a given set of humans want them to do. By default, language models optimize the next word prediction objective, which is only a proxy for what we want these models to do. Our results indicate that our techniques hold promise for making language models more helpful, truthful, and harmless. Making language models bigger does not inherently make them better at following a user's intent. For example, large language models can generate outputs that are untruthful, toxic, or simply not helpful to the user. In other words, these models are not aligned with their users." (My highlights).

"What sets ChatGPT apart from a simple chatbot is that it was specifically trained to understand the human intent in a question and provide helpful, truthful, and harmless answers. Because of that training, ChatGPT may challenge certain questions and discard parts of the question that don't make sense."

"Another research paper related to ChatGPT shows how they trained the AI to predict what humans preferred. The researchers noticed that the metrics used to rate the outputs of natural language processing AI resulted in machines that scored well on the metrics, but didn't align with what humans expected. The following is how the researchers explained the problem: "Many machine learning applications optimize simple metrics which are only rough proxies for what the designer intends. This can lead to problems, such as YouTube recommendations promoting click-bait."

"So the solution they designed was to create an AI that could output answers optimized to what humans preferred. To do that, they trained the AI using datasets of human comparisons between different answers so that the machine became better at predicting what humans judged to be satisfactory answers. The paper shares that training was done by summarizing Reddit posts and also tested on summarizing news. The research paper from February 2022 is called *Learning to Summarize from Human Feedback*."

"The researchers write: "In this work, we show that it is possible to significantly improve summary quality by training a model to optimize for human preferences. We collect a large, high-quality dataset of human comparisons between summaries, train a model to predict the human-preferred summary, and use that model as a reward function to fine-tune a summarization policy using reinforcement learning.""

So the developers have apparently done their best to avoid unintended negative consequences, which is extremely complex, but the responsible thing to do.

"What are the Limitations of ChatGPT?

"Quality of Answers Depends on Quality of Directions

"An important limitation of ChatGPT is that the quality of the output depends on the quality of the input. In other words, expert directions (prompts) generate better answers."

At least this is reassuring for education – it should be obvious when a learner or student has relied on ChatGPT only.

"Answers Are Not Always Correct

"Another limitation is that because it is trained to provide answers that feel right to humans, the answers can trick humans that the output is correct.

"Many users discovered that ChatGPT can provide incorrect answers, including some that are wildly incorrect. The moderators at the coding Q&A website Stack Overflow may have discovered an unintended consequence of answers that feel right to humans. Stack Overflow was flooded with user responses generated from ChatGPT that appeared to be correct, but a great many were wrong answers.

"The thousands of answers overwhelmed the volunteer moderator team, prompting the administrators to enact a ban against any users who post answers generated from ChatGPT. The flood of ChatGPT answers resulted in a post entitled: Temporary policy: ChatGPT is banned: "This is a temporary policy intended to slow down the influx of answers and other content created with ChatGPT. …The primary problem is that while the answers which ChatGPT produces have a high rate of being incorrect, they typically "look like" they "might" be good…"

The official announcement stated that OpenAI was eager to receive feedback about the mistakes:

"While we've made efforts to make the model refuse inappropriate requests, it will sometimes respond to harmful instructions or exhibit biased behavior. We're using the Moderation API to warn or block certain types of **unsafe content**, but we expect it to have some false negatives and positives for now. We're eager to collect user feedback to aid our ongoing work to improve this system."

"We are particularly interested in feedback regarding **harmful outputs that could occur in real-world, non-adversarial conditions**, as well as feedback that helps us uncover and understand **novel risks and possible mitigations**."

I do not know whether and to what extent these ChatGPT issues have been resolved. However, one would suppose some of such undesirable content would be unavoidable, given that the bot is trained on human interactions on social media.

Google's LaMDA

Google developed a similar chatbot, described here: LaMDA: our breakthrough conversation technology (blog.google) on May 18, 2021, by Eli Collins VP, Product Management and Zoubin Ghahramani, VP, Google Research. LaMDA is short for "Language Model for Dialogue Applications".

I do not know whether ChatGPT's truthfulness issues were resolved, but Google stresses this as a value that is built into their platform. LaMDA is also a conversational chatbot based on large language models and trained on big data.

They say:

"The long road to LaMDA

"LaMDA's conversational skills have been years in the making. Like many recent language models, including BERT and GPT-3, it's built on Transformer, a neural network architecture that Google Research invented and open-sourced in 2017. That architecture produces a model that can be trained to read many words (a sentence or paragraph, for example), pay attention to how those words relate to one another and then predict what words it thinks will come next."

Google was thus the forerunner in this space and OpenAI would have been fortunate to learn from its example..

"But unlike most other language models, LaMDA was trained on dialogue. During its training, it picked up on several of the nuances that distinguish open-ended conversation from other forms of language. One of those nuances is sensibleness. Basically: Does the response to a given conversational context make sense?

"LaMDA builds on earlier Google research, published in 2020, that showed Transformer-based language models trained on dialogue could learn to talk about virtually anything. Since then, we've also found that, once trained, LaMDA can be fine-tuned to significantly improve the sensibleness and specificity of its responses."

"Responsibility first

"These early results are encouraging, and we look forward to sharing more soon, but sensibleness and specificity aren't the only qualities we're looking for in models like LaMDA. We're also exploring dimensions like "interestingness," by assessing whether responses are insightful, unexpected or witty. Being Google, we also care a lot about **factuality** (that is, whether LaMDA sticks to facts,

something language models often struggle with), and are investigating ways to ensure LaMDA's responses aren't just compelling but correct. [My highlighting]

Google had already flagged this as a difficult but important challenge, among other challenges as described as follows:

"But the most important question we ask ourselves when it comes to our technologies is whether they adhere to our AI Principles [see here below]. Language might be one of humanity's greatest tools, but like all tools it can be misused. Models trained on language can propagate that misuse — for instance, by internalizing biases, mirroring hateful speech, or replicating misleading information. And even when the language it's trained on is carefully vetted, the model itself can still be put to ill use.

"Our highest priority, when creating technologies like LaMDA, is working to ensure we minimize such risks. We're deeply familiar with issues involved with machine learning models, such as unfair bias, as we've been researching and developing these technologies for many years. That's why we build and open-source resources that researchers can use to analyze models and the data on which they're trained; why we've scrutinized LaMDA at every step of its development; and why we'll continue to do so as we work to incorporate conversational abilities into more of our products."

Google published its Principles in June 2018, written by the CEO Sundar Pichai: AI at Google: our principles (blog.google):

"**At its heart, AI is computer programming that learns and adapts. It can't solve every problem, but its potential to improve our lives is profound.** At Google, we use AI to make products more useful—from email that's spam-free and easier to compose, to a digital assistant you can speak to naturally, to photos that pop the fun stuff out for you to enjoy.

"Beyond our products, we're using AI to help people tackle urgent problems. A pair of high school students are building AI-powered sensors to predict the risk of wildfires. Farmers are using it to monitor the health of their herds. Doctors are starting to use AI to help diagnose cancer and prevent blindness. These clear benefits are why Google invests heavily in AI research and development, and makes AI technologies widely available to others via our tools and open-source code.

"We recognize that such powerful technology raises equally powerful questions about its use. How AI is developed and used will have a significant impact on society for many years to come. As a leader in AI, we feel a deep responsibility to get this right. So today, we're announcing seven principles to guide our work going forward. These are not theoretical concepts; they are concrete standards that will actively govern our research and product development and will impact our business decisions.

"We acknowledge that this area is dynamic and evolving, and we will approach our work with humility, a commitment to internal and external engagement, and a willingness to adapt our approach as we learn over time."

"**Objectives for AI applications**

"We will assess AI applications in view of the following objectives. We believe that AI should:

1. "Be socially beneficial.

"**The expanded reach of new technologies increasingly touches society as a whole.** Advances in AI will have transformative impacts in a wide range of fields, including healthcare, security, energy,

transportation, manufacturing, and entertainment. As we consider potential development and uses of AI technologies, we will take into account a broad range of social and economic factors, and will proceed where we believe that the overall likely benefits substantially exceed the foreseeable risks and downsides.

"AI also enhances our ability to understand the meaning of content at scale. We will strive to make high-quality and accurate information readily available using AI, while continuing to respect cultural, social, and legal norms in the countries where we operate. And we will continue to thoughtfully evaluate when to make our technologies available on a non-commercial basis."

Google's heart seems to be in the right place, as they are mindful of the impact of their AI development on broader society. I cannot judge whether OpenAI adheres to the same values, but it does appear that Google takes the bigger picture seriously.

2. "Avoid creating or reinforcing unfair bias.

"AI algorithms and datasets can reflect, reinforce, or reduce unfair biases. We recognize that distinguishing fair from unfair biases is not always simple, and differs across cultures and societies. We will seek to avoid unjust impacts on people, particularly those related to sensitive characteristics such as race, ethnicity, gender, nationality, income, sexual orientation, ability, and political or religious belief."

3. "Be built and tested for safety.

"We will continue to develop and apply strong safety and security practices to avoid unintended results that create risks of harm. **We will design our AI systems to be appropriately cautious, and seek to develop them in accordance with best practices in AI safety research.** In appropriate cases, we will test AI technologies in constrained environments and monitor their operation after deployment."

4. "Be accountable to people.

"We will design AI systems that provide appropriate opportunities for feedback, relevant explanations, and appeal. Our AI technologies will be subject to appropriate human direction and control."

5. "Incorporate privacy design principles.

"We will incorporate our privacy principles in the development and use of our AI technologies. We will give opportunity for notice and consent, encourage architectures with privacy safeguards, and provide appropriate transparency and control over the use of data."

6. "Uphold high standards of scientific excellence.

"**Technological innovation is rooted in the scientific method and a commitment to open inquiry, intellectual rigor, integrity, and collaboration. AI tools have the potential to unlock new realms of scientific research and knowledge in critical domains like biology, chemistry, medicine, and environmental sciences. We aspire to high standards of scientific excellence as we work to progress AI development.**

"We will work with a range of stakeholders to promote thoughtful leadership in this area, drawing on scientifically rigorous and multidisciplinary approaches. And we will responsibly share AI knowledge by publishing educational materials, best practices, and research that enable more people to develop useful AI applications."

7. "Be made available for uses that accord with these principles.

"Many technologies have multiple uses. We will work to limit potentially harmful or abusive applications. As we develop and deploy AI technologies, we will evaluate likely uses in light of the following factors:

- *Primary purpose and use: …*
- *Nature and uniqueness: …*
- *Scale: (*whether the use of this technology will have significant impact)
- *Nature of Google's involvement: …"*

"AI applications we will not pursue

"In addition to the above objectives, we will not design or deploy AI in the following application areas:

1. "Technologies that cause or are likely to cause overall harm. Where there is a material risk of harm, we will proceed only where we believe that the benefits substantially outweigh the risks, and will incorporate appropriate safety constraints.
2. **"Weapons or other technologies whose principal purpose or implementation is to cause or directly facilitate injury to people.**
3. **"Technologies that gather or use information for surveillance violating internationally accepted norms.**
4. **"Technologies whose purpose contravenes widely accepted principles of international law and human rights."**

"AI for the long term

"While this is how we're choosing to approach AI, we understand there is room for many voices in this conversation. As AI technologies progress, we'll work with a range of stakeholders to promote thoughtful leadership in this area, drawing on scientifically rigorous and multidisciplinary approaches. And we will continue to share what we've learned to improve AI technologies and practices.

"We believe these principles are the right foundation for our company and the future development of AI. This approach is consistent with the values laid out in our original Founders' Letter back in 2004. There we made clear our intention to take a long-term perspective, even if it means making short-term tradeoffs. We said it then, and we believe it now."

The importance of being human

What are the implications of the existence of sophisticated platforms like ChatGPT for us humans and our productive lives?

NSTF held a discussion forum over two days last year on the **Creative economy, science and the 4IR**: Creative-economy-science-and-the-4IR-Media-Release.pdf (nstf.org.za) at which the importance of creativity and innovation were highlighted in the context of the Fourth Industrial Revolution (4IR). The era we find ourselves in (whatever we call it) is one in which all processes are being automated faster than we can keep up. My local bank branch used to inhabit an entire building, and a visit used to involve long queuing – now it is not much more than a counter, and only one or two clients are there when I pay it a rare visit. Many bank workers have had to find other jobs. This is part of the

enormous threat of this era – that millions of jobs will most certainly be lost due to computer automation and online processes.

In this context, it is important to assert and practice the best of what makes us human: art, music, poetry and other creative endeavours; as well as communication which is human-to-human. All these very human activities will still be required because *people* need to interact with *people*. No matter how well computer programmes can imitate humans in these activities, and beat humans at their own game, if there is not a live human behind an activity it cannot maintain its interest for other human beings. Where tasks need to be done that robots can do better than us, even I won't mind being served by one. But people still need people. *Motho ke motho ka batho babang.*

About communication

One learns when studying communication, that there are three aspects to communication: the message, the sender and the receiver. It is my understanding that the 'senders' must intend to communicate and to convey the message they formulate. Without an intentional (human) sender, and if someone does not take responsibility for a computer-generated message that is sent, no communication has taken place. The third aspect of communication is the 'receiver'. If the receiver does not understand the message and the intent of the sender, he or she has not received the message and again no communication has taken place. Mistakes in communication are frequent among humans, but the mistakes can be clarified, and the intent of the 'sender' is gleaned from the context so that the 'receiver' can understand the 'message' more clearly and even disregard any unintentional mistakes. In the robot (AI) environment, on the other hand,

To end on a somewhat philosophical note:

Can chatbots really be understood to be 'communicating', as senders of answers to their human questioners?

But then, what about other so-called vehicles of communication – can messages with false information 'forwarded many times' on social media be said to constitute true communication?

Where is the locus of truth?

FURTHER READING

A. **The OECD Framework for Classifying AI Systems to assess policy challenges and ensure international standards in AI:** by Jack Clark, Dewey Murdick, Karine Perset, Marko Grobelnik (February 17, 2022)

The OECD Framework for Classifying AI Systems to assess policy challenges and ensure international standards in AI - OECD.AI

AI Strategies and Policies in South Africa - OECD.AI

The OECD Artificial Intelligence (AI) Principles - OECD.AI

The OECD AI Principles promote use of AI that is innovative and trustworthy and that respects human rights and democratic values. Adopted in May 2019, they set standards for AI that are practical and flexible enough to stand the test of time:

1. Investing in AI research and development (OECD AI Principle) - OECD.AI

- *… to spur innovation in trustworthy AI that focus on challenging technical issues and on AI-related social, legal and ethical implications and policy issues.*
- *… to support an environment for AI research and development that is free of inappropriate bias and to improve interoperability and use of standards.*

2. Fostering a digital ecosystem for AI (OECD AI Principle) - OECD.AI

… for trustworthy AI. Such an ecosystem includes in particular digital technologies and infrastructure, and mechanisms for sharing AI knowledge, as appropriate. In this regard, governments should consider promoting mechanisms, such as data trusts, to support the safe, fair, legal and ethical sharing of data.

3. Shaping an enabling policy environment for AI (OECD AI Principle) - OECD.AI

- *… that supports an agile transition from the research and development stage to the deployment and operation stage for trustworthy AI systems. To this effect, they should consider using experimentation to provide a controlled environment in which AI systems can be tested, and scaled-up, as appropriate.*
- *… to encourage innovation and competition for trustworthy AI.*

4. Building human capacity and preparing for labour market transformation (OECD AI Principle) - OECD.AI

- *Governments should work closely with stakeholders to prepare for the transformation of the world of work and of society. They should empower people to effectively use and interact with AI systems across the breadth of applications, including by equipping them with the necessary skills.*
- *Governments should take steps, including through social dialogue, to ensure a fair transition for workers as AI is deployed, such as through training programmes along the working life, support for those affected by displacement, and access to new opportunities in the labour market.*
- *Governments should also work closely with stakeholders to promote the responsible use of AI at work, to enhance the safety of workers and the quality of jobs, to foster entrepreneurship and productivity, and aim to ensure that the benefits from AI are broadly and fairly shared.*

5. International co-operation for trustworthy AI (OECD AI Principle) - OECD.AI

- *Governments, including developing countries and with stakeholders, should actively cooperate to advance these principles and to progress on responsible stewardship of trustworthy AI.*
- *Governments should work together in the OECD and other global and regional fora to foster the sharing of AI knowledge, as appropriate. They should encourage international, cross-sectoral and open multi-stakeholder initiatives to garner long-term expertise on AI.*

- *Governments should promote the development of multi-stakeholder, consensus-driven global technical standards for interoperable and trustworthy AI.*
- *Governments should also encourage the development, and their own use, of internationally comparable metrics to measure AI research, development and deployment, and gather the evidence base to assess progress in the implementation of these principles.*

B. NSTF's discussion forum on the Creative economy, science and the 4IR

Creative economy, science and the 4IR (2-3 March 2022) – NSTF

Prof Tshilidzi Marwala (who was still the VC of the University of Johannesburg) addressed the NSTF discussion forum mentioned above (His speech can be read here: 22CreativeEconomySpeakerAddressProfMarwala.pdf (nstf.org.za)). He quoted the American writer, Prof Isaac Asimov (who wrote science fiction novels among others): "**It is only afterward that a new idea seems reasonable. To begin with, it usually seems unreasonable.** It seems the height of unreason to suppose the earth was round instead of flat, or that it moved instead of the sun, or that objects required a force to stop them when in motion, instead of a force to keep them moving, and so on…"

Prof Marwala says: **"I must emphasise that innovative thinking, the ability to come up with new ideas and intriguing solutions to problems, is a skill. In fact, it is this very skill that in many ways differentiates us from machines – for the moment at least."**

Among Marwala's 11 attributes required of a leader in the 21st century, he lists "4IR thinking", which includes looking to technology to solve the pressing challenges facing society. For example he says: "The most effective way to stop corruption is technologies of the 4IR: AI and blockchain." Among the 11 attributes he also lists "Computational thinking": "Computational machines increasingly make significant economic decisions in our banks, stock markets and government. To ensure that leaders are not reduced to speculating, they must adopt the ways of machines, which is computational thinking."

C. What is a chatbot?

Chatbots aim to serve. They are supposed to gather from the words in your request what it is that you need help with. If they can do that, they tend to be very useful. IBM's website explains their use as follows:

What is a chatbot? | IBM

"A chatbot is a computer program that uses artificial intelligence (AI) and natural language processing (NLP) to understand customer questions and automate responses to them, simulating human conversation.

"…The longer an AI chatbot has been in operation, the stronger its responses become. So an AI chatbot using deep learning may provide a more detailed and accurate response to a query, and especially to the intentions behind the query, than a chatbot with recently integrated algorithm-based knowledge.

"Today's AI chatbots use natural language understanding (NLU) to discern the user's need. Then they use advanced AI tools to determine what the user is trying to accomplish. These technologies rely on machine learning and deep learning—elements of AI, with some nuanced differences—to

develop an increasingly granular knowledge base of questions and responses that are based on user interactions. This improves their ability to predict user needs accurately and respond correctly over time.

“Business use is equally varied. Marketers use AI chatbots to personalize customer experiences, IT teams use them to enable self-service, and customer contact centers rely on chatbots to streamline incoming communications and direct customers to resources.

“Today, chatbots can consistently manage customer interactions 24x7 while continuously improving the quality of the responses and keeping costs down. Chatbots automate workflows and free up employees from repetitive tasks. A chatbot can also eliminate long wait times for phone-based customer support, or even longer wait times for email, chat and web-based support, because they are available immediately to any number of users at once. That’s a great user experience—and satisfied customers are more likely to exhibit brand loyalty.”

The opinions expressed above are those of Ms Jansie Niehaus, and do not necessarily reflect the views of the Executive Committee or members of the NSTF.