

BENEFITS AND CONCERNS

of Using Emerging Artificial Intelligence Chatbots With Work in NDT

BY JOHN C. ALDRIN

While most of the papers in this special issue explore the use of artificial intelligence and machine learning (AI/ML) to support the evaluation of nondestructive testing (NDT) data and assist with the classification of NDT indications, there are other important ways that emerging AI tools may impact how we work in NDT. The article discusses the recent emergence of AI chatbots, also referred to as generative artificial intelligence agents or large language models (LLMs), and highlights the potential benefits and risks as part of work in the NDT field.

Introduction

ChatGPT, launched in late 2022, has been getting a lot of attention due to its ability to generate human-like text responses to various tasks such as answering questions or generating summaries of content (Molla 2023). For many of us who have been working with computer algorithms since the 1990s, the idea of developing computer-assisted tools, like

expert systems, has been around for a long time. What makes these systems different today? Recent breakthroughs in the AI field are enabling such tools to be much more useful than past generations (Wolfram 2023). To start, these models are very large. The GPT-3 architecture is from OpenAI, and the language model ChatGPT has been reportedly trained on around 45 terabytes of text data,



which is roughly one quarter the size of the Library of Congress. Second, these models also operate differently than models of the past. ChatGPT and other LLMs are trying to produce a reasonable continuation of the text that was initially provided. Some people refer to this as a “next word guessing engine.” Based on input phrases, these LLMs are essentially evaluating the probability of what the next word should be. What is interesting is the latest generation of models don’t always choose the highest-ranked word. Because some randomness is involved in the model response, these AI chatbots appear to produce much more variety and interesting content. For more details on how ChatGPT works, see Wolfram (2023).

Potential Benefits

Let’s ask ChatGPT about how it might help NDT professionals with their job. (Note: my prompts are in blue and ChatGPT’s responses are in red.)

ChatGPT Q&A Part I

What are the benefits or positives of using emerging AI chatbots like ChatGPT with work in nondestructive testing?

There are several potential benefits to using emerging AI chatbots like ChatGPT in nondestructive testing (NDT) applications. Some of these benefits include:

- 1. Improved speed and efficiency: AI chatbots can perform tasks much faster than humans, and can work around the clock without breaks. This means that NDT tasks can be completed more quickly and efficiently, which can lead to cost savings and improved productivity.**
- 2. Increased accuracy: AI chatbots can be trained to perform NDT tasks with a high degree of accuracy, which can reduce the risk of errors and improve the reliability of test results.**
- 3. Enhanced safety: NDT tasks can often be dangerous or involve exposure to hazardous materials. By using AI chatbots to perform these tasks, human workers can be kept out of harm’s way, reducing the risk of accidents and injuries.**
- 4. Scalability: AI chatbots can be easily scaled up or down depending on the needs of the project, making it easier to handle large volumes of work or adjust to changes in demand.**
- 5. Better data analysis: AI chatbots can process large amounts of data quickly and accurately, allowing for more detailed analysis of test results and better insights into the health of the materials being tested.**

Overall, the use of emerging AI chatbots like ChatGPT has the potential to significantly improve NDT processes, making them faster, more accurate, and safer for workers.

This reads pretty well. I must admit I’m skeptical of some of these points, but I do feel there are clear opportunities for AI chatbots to assist humans to be more efficient with some laborious tasks.

Jill Ross, the ASNT Director of Publications, wrote about the potential impact of ChatGPT on editing in the February 2023 issue of *Materials Evaluation*, using ChatGPT: “AI has the potential to revolutionize the editing and publishing industry by automating various tasks that are traditionally done by human editors and publishers” (Ross 2023). For example, one application is writing summaries and abstracts. AI can take a longer piece of writing and summarize it into a shorter version in seconds, which can be particularly useful for editors and publishers who need to quickly grasp the main ideas of a piece of writing. Another one of the most promising aspects of the latest AI agents is improved language editing for non-English speakers (Kim 2023). Such tools, currently free to everyone around the world, can produce generally good-quality, edited paragraphs in seconds. Coupling this feature with existing AI-based text translation tools, the quality and accessibility of published materials is being positively impacted throughout the world.

Taking this one step further, Ripi Singh has been one of the leading early investigators of the potential of collaborative writing with AI agents. In 2021, Singh and Garg created their own interface script to engage with GPT-3 and collaborated with it to research human factors in decision-making for NDE 4.0. The result was published in the *Journal of Nondestructive Evaluation* (Singh et al. 2021) and was the first NDT paper having an AI agent as a co-author. The article and a more recent blog post (Vrana et al. 2023) provide evidence of the opportunities for practical human-machine co-working at the linguistic and cognitive levels. (Also, see the NDE Outlook article in this issue on page 17 [Singh and Garg 2023].)

Beyond just technical writing, AI agents like ChatGPT can also be a valuable tool for programmers in several practical ways (Timothy 2023). ChatGPT is best used as a companion tool, generating efficient algorithms for specific tasks, answering coding questions, and providing help with ideas. It can also translate code to another language and in certain instances help debug your code. Tyler Lesthaeghe at the University of Dayton Research Institute has been experimenting with the potential

use of AI tools to accelerate NDT research. “I presented ChatGPT with a request to develop a Python script that would tackle an old challenge in automated geometry generation for finite element modeling. It generated a script that, with some minor tweaks, was functional; but it, unfortunately, did not actually solve the fundamental problem. However, it did make me aware of a library that I was previously unfamiliar. I later tested ChatGPT with a more general task—plot the dispersion curves for a generic isotropic material. It set up the script reasonably well, produced material property values for the material I requested, and generated functional code producing the correct kind of plot. However, the actual calculation, and thus the resulting curves themselves, were complete nonsense. ChatGPT can, however, be ‘coached.’ So, with a knowledge of how to properly solve the problem, you can provide it with that explanation, and it will ultimately generate the right output. While I think there is real potential value in accelerating the more trivial parts of code development for NDE applications, and it may be useful in providing new perspectives or potential solution paths for more challenging kinds of problems, it performs best with a human in the loop.”

Risk and Concerns

Let’s now ask ChatGPT about risks and concerns with NDT professionals using it.

ChatGPT Q&A Part II

What are the risks or potential issues using emerging AI chatbots with work in nondestructive testing?

While there are potential benefits to using emerging AI chatbots like ChatGPT in nondestructive testing (NDT), there are also some risks and potential issues that need to be considered. Some of these include:

1. **Lack of human judgment:** AI chatbots may not have the same level of judgment or intuition as a human worker, which could lead to errors or misinterpretations of data.
2. **Limited flexibility:** AI chatbots are only able to perform tasks that they have been specifically programmed to do, which could limit their usefulness in certain NDT applications.
3. **Data bias:** AI chatbots may be trained on biased or incomplete data, which could lead to inaccurate or unreliable test results.
4. **Cybersecurity risks:** AI chatbots are vulnerable to cybersecurity threats like hacking or data breaches, which could compromise sensitive NDT data.

5. **Ethical concerns:** The use of AI chatbots in NDT raises ethical questions about the role of technology in the workplace, and the potential impact on human workers.

It is important to carefully evaluate these risks and potential issues when considering the use of AI chatbots in NDT applications, and to take steps to mitigate these risks through careful planning, rigorous testing, and ongoing monitoring and maintenance.

What sources did you use to provide these responses on the benefits and risk for using AI for NDT?

As an AI language model, I have been trained on a large corpus of text data that includes a wide range of sources on the topic of artificial intelligence and nondestructive testing. Some of the sources that may have contributed to my response include:

1. Research articles and publications from academic journals, conferences, and industry experts on AI and NDT.
2. Online articles and reports from reputable sources such as industry publications, news outlets, and research organizations.
3. White papers and case studies from companies and organizations that are involved in developing or using AI for NDT applications.
4. Technical documentation and specifications for AI technologies and tools that are relevant to NDT.

It is important to note that my responses are generated based on patterns and associations in the data I have been trained on, and may not always reflect the most current or accurate information on the topic. It is always a good idea to consult multiple sources and consult with subject matter experts when making decisions about the use of AI in NDT.

Humans are required to cite their specific sources when writing. It is a common courtesy to recognize one’s peers. But more important, citation of sources is a critical part of referencing where ideas come from and validating the foundations of their work. Can you please provide articles you are leveraging to respond to my questions on artificial intelligence and nondestructive testing above?

As an AI language model, I don’t have access to a list of specific articles or sources that have contributed to my response. This is because my responses are generated based on patterns and associations in the data I have been trained on, which includes a vast amount of text data from a wide range of sources.

However, I can suggest some relevant articles and sources related to artificial intelligence and nondestructive testing that you may find useful:

1. “Applications of artificial intelligence and machine learning in non-destructive testing” by A. T. Arulrajah, P. R. W. Nadeeka, and M. A. Ali. This is an academic research paper that provides an overview of the applications of AI and machine learning in NDT.
2. “Non-Destructive Testing and Inspection Market - Growth, Trends, COVID-19 Impact, and Forecasts (2021 - 2026)” by Mordor Intelligence. This is a market research report that provides insights into the trends and drivers in the NDT market, including the role of AI and machine learning.
3. “Artificial Intelligence in Nondestructive Testing: A Review” by M. Elsafi, S. Saeed, and S. M. Zubair. This is another academic research paper that provides a comprehensive review of the use of AI in NDT, including the benefits and challenges.
4. “Nondestructive Testing Using Artificial Intelligence: Trends and Opportunities” by S. K. Tiwari and R. K. Jain. This is a research article that discusses the use of AI in NDT, with a focus on the potential benefits and challenges.

I hope you find these resources helpful in your research on AI and NDT.

At first, this series of responses appears appropriate and honestly impressive. However, there is one major issue if you dig deeper; none of these suggested references in the last response are actually real publications. They are all fabrications.

Why does this happen? GPT-3 is a language model that uses deep neural networks to predict the next word in a sequence of text based on the context provided by the previous words. Sometimes, when the predicted prompt responses are not precisely matched, the LLM will still do its best to string together the most plausible text response. Unfortunately, for some instances where precise responses are expected, sometimes they turn out to be nonsense.

Besides fake references, poor logic and unexpected responses are an issue for AI agents in 2023. The term *hallucinations* has been given to such fabricated responses (Alkaissi and McFarlane 2023). This issue becomes more critical when such tools are applied to technical and scientific work, like NDT, where specific details are of the utmost importance and where we can't afford to have such errors and fabrications. On NDT.net, there is a thread highlighting the bad idea of “ChatGPT3 writing your inspection procedure” (Bisle 2023). Clearly, AI agents are not ready to be given large complicated

technical writing tasks and be expected to produce error-free content. From my perspective, this is OK. NDT technicians, engineers, and researchers should be leading and responsible for the quality of written procedures, reports, and scientific publications.

The other issue that the aforementioned interchange tried to highlight is the lack of any means to reference and verify where such content originated. I've written about the benefits and risks of AI for NDT in the past and was curious where ChatGPT was getting its material. While it is impressive that such AI agents can generate articulate responses to such questions, I do see an ethical issue. If these language models are being trained using material on the order of the content of the Library of Congress, shouldn't they do a better job of providing the source material for their response? To some degree, the current versions of these AI tools operate like efficient plagiarism agents, which is the antithesis of quality technical and scientific writing that depends on collegial citation.

The Future

These tools have come a long way in recent years and will only get better. While ChatGPT is based on GPT-3, OpenAI recently released GPT-4, which has received many positive reviews (Metz and Collins 2023). (While there is a monthly charge to access GPT-4 directly, Microsoft Bing Chat does provide free limited access to GPT-4 today.) There are also a number of other promising AI tools to explore today like Google's Bard, DeepMind Sparrow, and Amazon Titan. In terms of knowledge capability, GPT-4 has been trained to be more precise and OpenAI claims it can score a 1300 (out of 1600) on the SAT. So, training on a wider depth of material and taking more care with the content selection will help. But, to some degree, if these AI agents are trained using the broad history of human writing, all of the positives and the negatives of our writing will be baked into these algorithms. The current black-box architecture will make it challenging to eliminate false or offensive responses.

Going forward, the most effective way of using such tools will be in a collaborative way. This will follow our general experience with the application of AI/ML for evaluating NDT data, where maintaining a critical role for human inspectors ensures NDT data quality and helps compensate for instances of poor AI performance. (See Lindgren 2023 on page 35 in this issue for more discussion on this topic.) Workers are already finding ways to leverage these tools effectively while doing their job. In a recent survey, over 40% of Americans said they were using generative AI technology at work

(Molla 2023). While new technologies certainly can cause disruptions, they may ultimately lead to more and better-quality work, much like the impact of the personal computer or the internet. University instruction is already striving to rethink how to integrate such tools into their curriculum and promote best practices (Yang 2023). It is critical to understand how to create appropriate prompts for getting the best information, while also understanding the risk and quality issues of the output.


One of the biggest issues going forward concerns plagiarism, copyright concerns for human content providers, and how this technology could be better regulated. Artists and writers are beginning to take action to defend their intellectual property from so-called “fair use” (DeSignore 2023). Daniel Gervais, a professor at Vanderbilt Law School who specializes in intellectual property law states that it hinges on the following: “What’s the purpose or nature of the use and what’s the impact on the market” (DeSignore 2023). Basically, it comes down to how you are using the output. Is it for research or commercial purposes? If commercial, one needs to be extremely careful. These questions and concerns are going to greatly impact the future of this technology, and how widely and rapidly it will be used.

The regulation of AI is expected to evolve rapidly and must address the safe application of this technology. To date, regulation is being led by the EU and China, while the US response has been fairly limited in scope. The White House’s *Blueprint for an AI Bill of Rights* highlights the need for better decision-making including explanations: “Automated systems should provide explanations that are technically valid, meaningful, and useful to you and to any operators or others who need to understand the system, and calibrated to the level of risk based on the context” (Klein 2023). But experts generally agree we have made almost no progress on explaining what is happening inside these LLMs (Klein 2023). There is a clear need to be able to comprehensively validate AI performance, but this appears to be greatly complicated by how complex these algorithms have become. Work on Explainable AI—a set of tools that help one understand and interpret the outputs generated by ML algorithms—is progressing, but it will take time to get there.

One consideration for our community: What if we created our own NDT Chatbot, let’s say residing behind the ASNT login, trained using ASNT-copyrighted materials, for example back issues of *Materials Evaluation*, and maybe even handbooks? Based on what GPT-4 is doing, it is clear such a tool could pass a Level III exam. If done right, this could be a valuable resource for the community. Of

course, we’d have to first ensure that the answers are consistently correct, just as we have reviewers ensure our handbooks and publications are as error-free as possible. I feel the technology would also need to produce the source(s) for its answer to the user, so we have a record to check and verify that the answer is correct. If poor responses are discovered, we must also have the means of correcting it.

While we can imagine all of the positive uses for such AI agents, they can just as easily be deployed for nefarious causes today. For example, these tools will likely improve the social engineering that is being used to fleece people of personal information and money through predatory emails, robo-calls, and social media. It is critical to consider the trade-offs of organizing our body of knowledge into one easily accessible place. Ripi Singh has some very important insight on this going forward: “The ‘Vulnerable World Hypothesis’ is a topic that deserves our undivided attention at every ASNT conference as a single body of professionals committed to Creating a Safer World!” We can start with Generative AI as the first item on the list to be addressed, now” (Singh and Garg 2023).

While I don’t have all the answers and definitely can’t predict the future, I do want to encourage more discussion and feedback on this important topic within ASNT. This topic has been brought up in the ASNT AI/ML Committee recently and we plan to explore possible guidance for the use of generative AI in NDT going forward. (As well, please feel free to share your thoughts with me at aldrin@computationaltools.com or get involved with the ASNT AI/ML Committee.) 

AUTHOR

John C. Aldrin: Computational Tools, Gurnee, Illinois 60031, USA; aldrin@computationaltools.com

CITATION

Materials Evaluation 81 (7): 28-34
<https://doi.org/10.32548/2023.me-04361>
©2023 American Society for Nondestructive Testing

REFERENCES

- Alkaissi, H., and S. I. McFarlane. 2023. “Artificial hallucinations in ChatGPT: Implications in scientific writing.” *Cureus* 15 (2). <https://doi.org/10.7759/cureus.35179>.
- Bisle, W. 2023. “ChatGPT3 writing your inspection procedure?” NDT.net forum. 29 January 2023. <https://www.ndt.net/forum/thread.php?msgID=84722>.
- DeSignore, P. 2023. “AI and the copyright problem: Making sense of generative AI copyright issues.” *Medium*. 4 March 2023. <https://medium.com/geekculture/ai-and-the-copyright-problem-97da479a9ccd>.
- Kim, S.-G. 2023. “Using ChatGPT for language editing in scientific articles.” *Maxillofacial Plastic and Reconstructive Surgery* 45. <https://doi.org/10.1186/s40902-023-00381-x>.

- Klein, E. 2023. "The surprising thing A.I. engineers will tell you if you let them." *The New York Times*. 16 April 2023. <https://www.nytimes.com/2023/04/16/opinion/this-is-too-important-to-leave-to-microsoft-google-and-facebook.html>.
- Lindgren, E. 2023. "Validated and deployable AI/ML for NDT data diagnostics." *Materials Evaluation* 81 (7): 35-42.
- Metz, C., and K. Collins. 2023. "10 ways GPT-4 is impressive but still flawed." *The New York Times*. 14 March 2023. <https://www.nytimes.com/2023/03/14/technology/openai-new-gpt4.html>.
- Molla, R. 2023. "AI is flooding the workplace, and workers love it." *Vox*. 10 April 2023. <https://www.vox.com/technology/23673018/generative-ai-chatgpt-bing-bard-work-jobs>.
- Ross, J. 2023. "AI and the future of publishing." *Materials Evaluation* 81 (2): 7.
- Singh, R., and V. Garg. 2023. "Nurturing generative AI: balancing innovation and responsibility." *Materials Evaluation* 81 (7): 17-20.
- Singh, R., V. Garg, and GPT-3. 2021. "Human factors in NDE 4.0 development decisions." *Journal of Nondestructive Evaluation* 40. <https://doi.org/10.1007/s10921-021-00808-3>.
- Timothy, M. 2023. "9 practical uses of ChatGPT in programming." *MUO*. 20 March 2023. <https://www.makeuseof.com/chatgpt-programming-practical-uses/>.
- Vrana, J., R. Singh, ChatGPT. 2023. "This is ChatGPT; how may I help you?" *Materials Evaluation*, 81 (2) 17-18.
- Wolfram, S. 2023. "What is ChatGPT doing...and why does it work?" *Stephen Wolfram Writings*. 14 February 2023. <https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/>.
- Yang, H. 2023. "How I use ChatGPT responsibly in my teaching." *Nature*. 12 April 2023. <https://doi.org/10.1038/d41586-023-01026-9>.

VALIDATED AND DEPLOYABLE AI/ML FOR NDT DATA DIAGNOSTICS

BY ERIC LINDGREN

While artificial intelligence/machine learning (AI/ML) methods have shown promise for the analysis of image and signal data, applications using nondestructive testing (NDT) for managing the safety of systems must meet a high level of quantified capability. Engineering decisions require technique validation with statistical bounds on performance to enable integration into critical analyses, such as life management and risk analysis. The Air Force Research Laboratory (AFRL) has pursued several projects to apply a hybrid approach that integrates AI/ML methods with heuristic and model-based algorithms to assist inspectors in accomplishing complex NDT evaluations. Three such examples are described in this article, including a method that was validated through a probability of detection (POD) study and deployed by the Department of the Air Force (DAF) in 2004 (Lindgren et al. 2005). Key lessons learned include the importance of considering the wide variability present in NDT applications upfront and maintaining a critical role for human inspectors to ensure NDT data quality and address outlier indications.

Introduction

There is a growing increase in interest and attention in AI/ML, which are statistical methods for data analysis. The promise of AI/ML is to use statistical methods to self-extract attributes in the data, such as relationships and/or trends in data that are not as quickly and reliably made through typical human observation. The DAF has embraced the use of these tools for applications where it can accelerate decision-making in representative campaigns, as shown in Figure 1. The objective defined for one of these efforts is summarized as: “The Air Force aims to harness and wield the most optimal forms of artificial intelligence to accomplish all mission-sets of the service with greater speed and accuracy” (USAF n.d.).

With the potential to secure more NDT data through the transformation to fully digital instruments connected as envisioned by the Internet of Things (IoT) and NDE 4.0, there is an increased interest to use AI/ML methods as the diagnostic tool to determine if a flaw is present in NDT data. Justification for the use of AI/ML includes improved accuracy, improved reliability, and faster disposition time by decreasing or eliminating dependence on human interpretation and analysis of NDT data. The initial focus for the use of AI/ML addresses the detection of flaw indications, although there is exploration in the use of AI/ML to provide additional information on characterizing the size and location of discontinuities.

When considering the applicability of AI/ML for flaw detection, it is important to recall that these technical approaches are based on statistical methods, namely regression or classification of data. The concept includes the use of multiple statistical methods in parallel combined with multiple layers of analysis to extract statistical trends in the data to enable decisions that are not readily detectable through more classical methods. These multidimensional data analysis methods frequently are called neural networks. These approaches can either be trained using data with known ground truths called supervised AI/ML, or be allowed to form the statistical relationships without training data, called unsupervised AI/ML. As these methods rely on

Figure 1. The Department of the Air Force artificial intelligence/machine learning campaign illustration.



US AIR FORCE GRAPHIC BY TRAVIS BURCHAM