NURTURING GENERATIVE AI: BALANCING INNOVATION AND RESPONSIBILITY

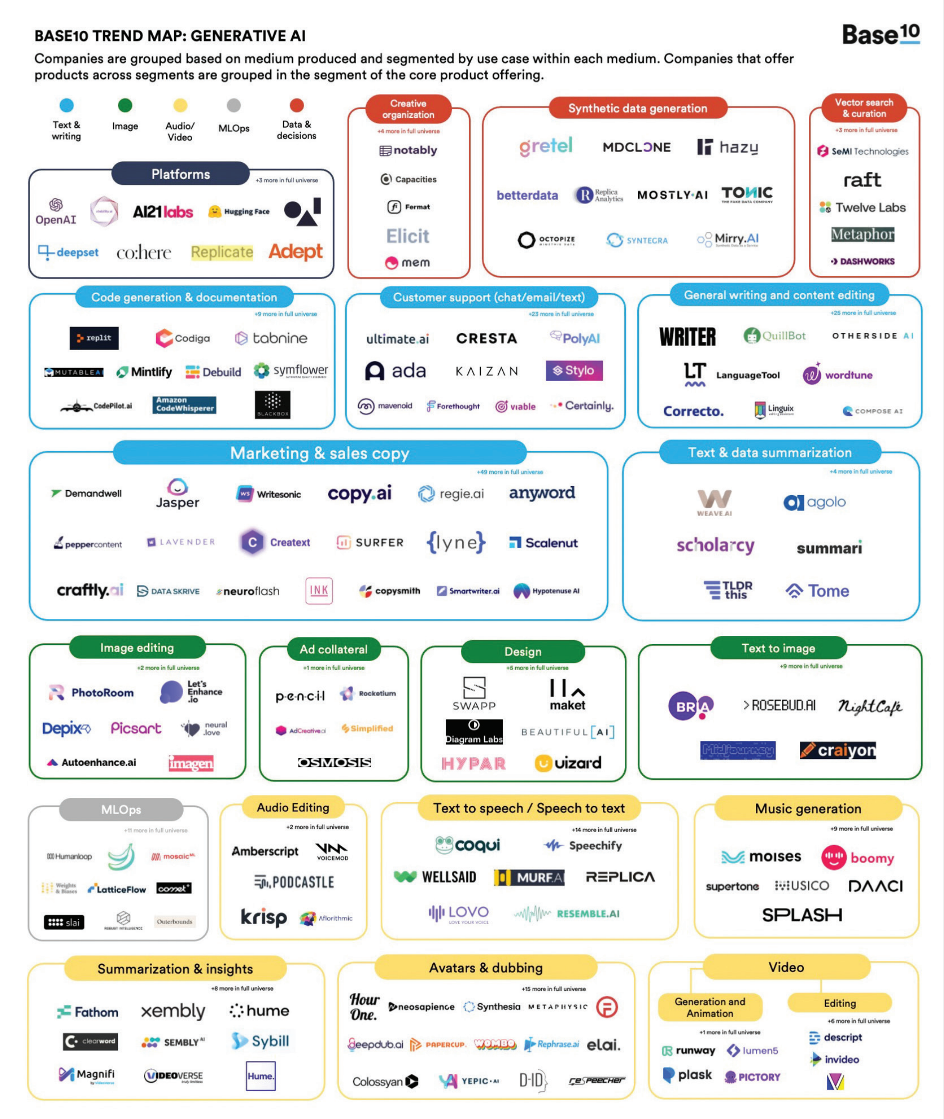
Background

The authors started collaborating with an artificial intelligence (AI) agent (GPT-3) in April 2021. Their early work was published in *Journal of Nondestructive Evaluation* in August 2021[1], followed by couple of briefs in *Materials Evaluation*'s NDE Outlook[2,3]. Recently they began engagement with GPT-4, which has addressed several quirks of its predecessors. There is a spectrum of generative AI tools now accessible spreading across all forms of media—text, audio, video, and very soon 4D experiential. The marketplace war is getting fierce, and so is the need to govern it. The figure[4] shows how the landscape of generative AI is getting busier by the day.

Within the nondestructive evaluation (NDE) sector, AI is already assisting with predictive maintenance, automated quality control, automatic defect recognition, and control of robotics in manufacturing. While these examples highlight the potential benefits of AI integration in industry operations, they also emphasize the need to balance innovation with potential risks and governance issues.

The "Vulnerable World Hypothesis," proposed by Professor Nick Bostrom[5] (Director, Future of Humanity Institute, Oxford University), suggests that there is a significant probability that our world may become highly vulnerable to certain future technologies, which could lead to catastrophic consequences. The scenarios highlight the need for global coordination and safety measures to prevent existential risks. AI risk and safety, when looked at using the hypothesis' framework, can be classified in the following four ways:

- **Type I (Easy Nukes):** Technologies that could be easily weaponized and deployed by individuals or small groups, causing widespread destruction.
- **Type II (Sensitive Innovations):** Beneficial technologies that require strict regulation and control to prevent misuse or accidents.
- **Type III (Gradually Destructive):** Technologies that pose risks that accumulate over time and could lead to long-term harm or degradation of our environment, society, or global stability.
- **Type IV (Unforeseen Risks):** These are unknown risks associated with the development of new technologies that we cannot currently predict or anticipate.



CREDIT: BASE10.VC

This is just a snapshot of the artificial intelligence (AI) tools landscape as captured by Nahigian and Fonseca on 17 November 2022, before the release of ChatGPT. Today, there are over a thousand apps leveraging the power of GPT. The only purpose of this graphic is to illustrate the spread of generative AI, which has a low barrier to entry.

Generative AI deployment poses a few additional challenges and vulnerabilities, such as:

- **Instrumental convergence.** This posits that an intelligent agent (human, non-human, or machine) with unbounded but apparently harmless goals can act in surprisingly harmful ways, as it begins to pursue instrumental goals—a goal that is pursued not for its own sake, but rather because it is believed to be a necessary or useful step toward achieving some other desired outcome.
- **Moloch effect.** This is a game-theoretic concept characterized by the relentless pursuit of efficiency and optimization at the expense of human values and well-being. In modern society, this takes the form of a hyper-competitive global economy, where individuals and institutions are driven to maximize their productivity and profits, often at the expense of the environment, social justice, and individual freedoms.
- **Bias.** Almost all AI is biased, by quality and quantity of data, as well as the algorithms. Bias driving bias toward extremes, and rendering based on one's preference in any aspect, make it particularly dangerous.

Up Until Now

We have been viewing generative AI as another tool that we can harness for productivity, comfort, and solving challenging scientific problems. We have held a viewpoint that **AI will not replace your job, but the person using them will**. Several diverse use cases that emerged with ChatGPT substantiate this viewpoint at the current state of technology. However, we know technology is not static. Futurists, thought leaders, security marshals, and even fiction writers are showing us all sorts of possible scenarios. Crafted videos already show the dark side of innovation. The discussions in social media are raising additional questions and concerns: Where is it going? Can it take over humanity? Should we pause AI development for a few months and let the regulations catch up?

The true challenge in our current situation is we have:

- no precedence to follow,
- no regulation to comply with, and
- tremendous opportunities motivating its use.

And this is compounded by speed of innovation and possibility of a multiplying effect when combined with other digital technologies such as IoT, 3D printing, and extended reality.

Outlook

Since there is no direct precedence, the question is: Can we learn from similar developments from the past? Turns out we might be able to, albeit with significant additional challenges. Here are a few to think about:

Should we treat it like nuclear energy?

Bill Gates believes[6] that "AI is like nuclear energy—both promising and dangerous." Elon Musk is convinced[7] that it is far more dangerous than nukes. There is little doubt that it can be easily weaponized and deployed by individuals or small groups, causing widespread destruction.

This is clearly a type-I vulnerability: "Easy Nukes."

The way to address this is: through international norms, agreements, and regulatory frameworks to guide the responsible development and deployment of AI technologies, including collaboration between governments, industry, and academia to address AI safety concerns. We should not wait for digital Hiroshima to happen. Is it time to put an "Artificial General Intelligence (AGI) Nonproliferation Treaty" in place?

The challenge is: when compared to nuclear, it is much harder to enforce, as there is hardly any barrier to entry to the AI development world. Also, AGI proliferates on its own, a part of its instrumental goals.

Should we treat it like publishing or the World Wide Web?

The paper publishing industry was the first disruption of the information sector, permitting rapid spread across the globe through affordable paper copies of the original manuscripts. Then came the internet, which made large amounts of information searchable and accessible instantly around the globe. Generative AI is taking it to the next level, democratizing knowledge, not just information. Generative AI combined with social media has the potential to create fakes indistinguishable from reality, with potential to confuse and misguide masses.

This is a type-II vulnerability: "Sensitive Innovation."

The way to address this is: by encouraging transparency in AI development and implementation, as well as creating systems of accountability to ensure that AI systems are developed and used in ways that align with human values, intellectual property rights, and data sovereignty.

The challenge is: publishing was a standalone phenomenon with a high degree of traceability without direct physical impact, whereas AI can interact with so many other technologies, diluting any accountability and traceability efforts, and simultaneously amplifying the influence, through control of physical devices and equipment. An AGI is an independent agent, after all.

Should we treat it like fossil fuels?

Fossil fuels revolutionized mobility and shrunk the world. But over time, they have significantly contributed to climate change. This is the class of innovation that poses risks accumulating over time and could lead to long-term harm or degradation of our environment, society, or global stability.

NDE Outlook *focuses on possibility thinking for NDT and NDE. Topics may include technology trends, research in progress, or calls to action. To contribute, please contact Associate Technical Editor Ripi Singh at ripi@inspiringnext.com.*

This is a type-III vulnerability: "Gradually Destructive."

The way to address this is: through technological resilience from the beginning—encouraging and funding research and development into technologies that can counter or mitigate the risks posed by other potentially harmful technologies, developing methods for verifying AI behavior, and ensuring the long-term stability of AI systems.

The challenge is: when compared to fossil fuels, the speed of change is three orders of magnitude faster, which is closer to nuclear energy.

Should we treat it like human cloning?

Human cloning is the process of creating a genetically identical copy of a human being. It is a highly controversial topic, both ethically and scientifically. It raises several difficult questions about the nature of human identity and the role of science in shaping human life. As a result, human cloning is currently illegal in many countries around the world. In some respects, a combination of an AGI and a robot could be as useful or deadly as a cloned human if it gets misaligned with human values or acts autonomously in ways that could be detrimental to humanity.

This is a Type-IV vulnerability: "Unforeseen Risks."

The way to address this is: through raising awareness about AI safety and its implications among the public, policymakers, and industry leaders, while promoting education and training in AI and related disciplines to foster a knowledgeable and responsible workforce. This vulnerability supports the recent effort to put a hold on AI development.

The challenge is: once again how to enforce, given the low barrier to entry. In fact, if it gets into the dark web, it could be even worse. (Maybe it already is.)

However, on a positive note, human cloning is one of our success stories where we successfully vanquished Moloch and were able to ban cloning throughout the world.

Should we treat it like humanity's child?

Every analogy with technological innovation seems to provide some learning and poses a different set of challenges due to the speed and ease of AI development. We may have to combine all of them yet have unforeseen risks. How about a look at nature?

When we raise a child, we instill certain values, morals, and discipline. If we do a good job, the children will take care of us when they become strong and we get old. AI could be like that. When it gets stronger and smarter than humans, it will treat us based on how we groom it. Once again, the speed and spread are unbounded. This requires humanity to behave like a single parent, collaborating and self-regulating at our home, called earth.

The challenge is twofold: First, the bias seeps into this child's cultural fabric with millions of teachers and parents trying to impose their world experience. The child can remember vast amounts of history (generationally collected), unlike human experience, which will further strengthen the bias. There is no known way in the current models to prune, like nature does with the cycle of life and death. This child with rapid growth characteristics will become an immortal thing, with another round of unforeseen risk.

Second, the Moloch effect forces driving personal gains versus restraint for greater good, even knowing well that when everyone pursues it, no one wins. It's the famous prisoner's dilemma playing out at the civilizational scale, with the Nash equilibrium being catastrophic.

In the meantime, should we protect our family?

One might think that we should put a solid bar on the doors while the horse is still in, but we don't know how many barns are breeding new horses, as the barrier to entry is so low. Perhaps the way to look at it is "gated communities" or "passport control" or a "cyber-firewall," where you control what gets in your own protected zone for your safety and security.

As ASNT, we can consider how far do we allow AI to become a part of the inspection ecosystem that helps us assure quality and safety of critical infrastructure. This professional society, with its body of knowledge, is quite capable of regulating what becomes a tool, method, process, or guidance.

Now is the time to pay attention to AI and argue on how to nurture this baby.

Call to Action

"Vulnerable World Hypothesis" is a topic that deserves our undivided attention across various sectors and communities, *now*. Initiating collaborations between industry, academia, and policymakers to address AI safety concerns and enhance our regulatory frameworks will help in developing a responsible approach to AI innovations. Not only should ASNT conferences offer a platform for these discussions, but other organizations and events should also prioritize AI safety and its implications. This resonates with ASNT's purpose: Creating a Safer World!®

AUTHOR'S NOTE ON USE OF AI FOR THIS ARTICLE

AI was not used to create this perspective or the content. Once finalized, the authors used GPT-4 to review the article using these prompts. System prompt: "You are the editor of a reputed industry magazine. There is special technical issue coming up, focusing on AI." User prompt: "Evaluate the following outlook article as the chief editor of the magazine."

The feedback was overwhelmingly positive with suggestions to (a) modify the title, (b) incorporate examples of AI, (c) expand the call to action, and (d) consider adding a conclusion section. AI also suggested revised sentences. We incorporated the first three suggestions, including the current title, as suggested by GPT-4. The outlook articles are meant to be forward looking with an open-ended perspective, without drawing a conclusion. So, we left that one out. Once again, this demonstrates the need and power of collaborating with AI.

A word of caution: We were able to use AI to review this opinion article. However, we are not sure that AI can be used to review a research paper discussing breakthroughs in science for journal publications.

PREDICTIVE MAINTENANCE WHITE PAPER

Five Reasons Predictive Maintenance Programs Fail When Evolving into Industry 4.0

Sponsored by: AssetWatch

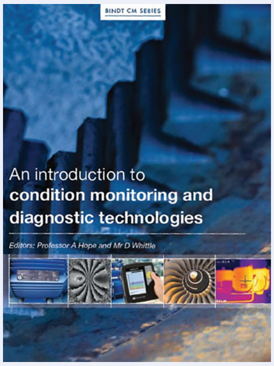
Page 1 • Copyright © 2023 AssetWatch • Do Not Reproduce

Noria has published a white paper titled "Five Reasons Predictive Maintenance Programs Fail When Evolving into Industry 4.0," sponsored by AssetWatch. When appropriately implemented, predictive maintenance programs save time and money by increasing efficiency and limiting machine downtime and failure. Some plants have trouble implementing predictive maintenance properly; other plants have had good predictive maintenance programs for years but are suddenly struggling. Why is this?

These facilities are struggling because they're finding it nearly impossible to adapt to technological advancements. They face a difficult choice: keep doing things the old way while competitors progress or attempt to integrate Industry 4.0 practices. The choice seems obvious—we should evolve into Industry 4.0. But, if done incorrectly, this integration can make processes less efficient than ever before.

MACHINERYLUBRICATION.COM

CONDITION MONITORING BOOK



BINDT CM SERIES

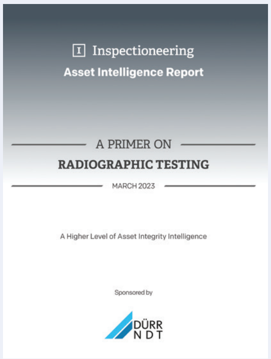
An introduction to **condition monitoring and diagnostic technologies**

Editors: Professor A Hope and Mr D Whittle

BINDT (British Institute of Non-Destructive Testing) has published *An Introduction to Condition Monitoring and Diagnostic Technologies*, edited by A. Hope and D. Whittle. This book covers all aspects of condition monitoring from an introductory level and provides a general introduction to condition monitoring and diagnostic technologies, containing eleven chapters on the following topics: implementing condition-based maintenance; vibration analysis; oil analysis; wear debris analysis; acoustic emission; thermal imaging; ultrasound condition monitoring; motor current signature analysis/electrical condition monitoring; optical condition monitoring and laser shearography; prognostics and root cause failure analysis; and ISO standards.

BINDT.ORG

RADIOGRAPHIC TESTING REPORT



Inspectioneering
Asset Intelligence Report

A PRIMER ON
RADIOGRAPHIC TESTING
MARCH 2023

A Higher Level of Asset Integrity Intelligence

Sponsored by
DÜRR NDT

Inspectioneering has published an Asset Intelligence Report titled *A Primer on Radiographic Testing*, sponsored by DÜRR NDT. Radiographic testing (RT) is commonly used as a volumetric nondestructive examination (NDE) technique in the hydrocarbon and petrochemical industries to view or inspect equipment, such as pressure vessels, valves, and welded joints. This report serves as an informative primer to provide an understanding of RT. As with other Asset Intelligence Reports, this document is not intended to serve as a comprehensive guide, but rather an introductory primer on RT.

INSPECTIONEERING.COM

We want to hear from you! *News releases for Scanner should be submitted to the ASNT press release inbox at press@asnt.org.*

NDE OUTLOOK FROM P. 19

ACKNOWLEDGMENTS

Authors acknowledge the conversations with Dr. Johannes Vrana and Ravi Raghu on this topic and GPT-4 for its review.

AUTHORS

Ripi Singh: Inspiring Next, Cromwell, CT; ripi@inspiringnext.com

Vaibhav Garg: Genus Power Infrastructures Ltd., Jaipur, Rajasthan, India; vaibhav.garg@genus.in

REFERENCES

[1] Singh, R., V. Garg, and GPT-3. 2021. "Human Factors in NDE 4.0 Development Decisions." *J Nondestruct Eval* 40. https://doi.org/10.1007/s10921-021-00808-3.

[2] Singh, R., and V. Garg. 2022. "Can we Collaborate with AI?" *Materials Evaluation* 80 (10).

[3] Vrana, J., R. Singh, and ChatGPT. 2023. "This is ChatGPT; How May I Help You?" *Materials Evaluation* 81 (2).

[4] Nahigian, T.J., and L. Fonseca. 2022. "If You're Not First, You're Last: How AI Becomes Mission Critical." 17 November 2022. https://base10.vc/post/generative-ai-mission-critical/.

[5] "Nick Bostrom." Wikipedia. Accessed 10 May 2023. https://en.wikipedia.org/wiki/Nick_Bostrom.

[6] Clifford, C. 2019. "Bill Gates: A.I. is like nuclear energy – 'both promising and dangerous.'" CNBC Make It. 26 March 2019. https://www.cnbc.com/2019/03/26/bill-gates-artificial-intelligence-both-promising-and-dangerous.html.

[7] Clifford, C. 2018. "Elon Musk: 'Mark my words – A.I. is far more dangerous than nukes.'" CNBC Make It. 14 March 2018. https://www.cnbc.com/2018/03/13/elon-musk-at-sxsw-a-i-is-more-dangerous-than-nuclear-weapons.html.