TECHNOLOGY, TRUTH AND TRUST

Illuminating a way forward for ethical AI

EXECUTIVE SUMMARY

This whitepaper explores ways in which creators and users of complex AI solutions can mitigate the many ethical risks posed to organizations and society, without undermining AI's profound ability to help solve our thorniest and most pressing challenges. As the pace of technological change increases, so does the importance of achieving a healthy balance between the awesome promise of AI advancements and mitigating the potential for existential threats they make apparent. In this way, AI epitomizes the term 'double-edged sword'.

Both edges of that sword were sharpened further in the wake of OpenAI's release of ChatGPT in November 2022. This major inflection point in the decades long arc of AI, sparked staggering increases in innovation, investment, and mainstream awareness, while simultaneously amplifying already loud alarm bells. The growing concerns are spread across a wide range of constituencies from corporations, workers, and citizens, to academia, governments, and regulators. Given how quickly things are moving, the concern is understandable. Can anyone be certain that AI will not someday exceed our capacity to understand or control it?

Once again, people's perception of AI cuts both ways. While 44% of executive leaders surveyed by KPMG in 2021 thought their industry was moving too fast on AI adoption, almost all those surveyed wished their own organization would move even faster[1]. If anything, this dichotomy has intensified further.

Based on decades of collective AI experience, and a deep understanding of our customers' risks and requirements, Nuix recognizes that what lies at the core of the AI dilemma is the concept of trust. Until recently, mistrust in AI has been the preserve of apocalyptic movies and novels. Our everyday encounters with AI were limited to routing our phone calls, correcting our spelling mistakes, or making movie recommendations. Yet, as we explore handing off more and more complex and substantive tasks to AI, scrutiny is quickly and rightly increasing around the credibility and accuracy of its output.

"The greatest source of all human security is trust". This was the incisive message at the heart of a powerful speech by Jeffrey Bleich, a Nuix Director, former US Ambassador to Australia and Special Counsel to President Obama. Bleich's address to business executives in April 2023 emphasized the important interconnectedness between technology, truth, and trust in our society. For AI to be given such a potent role in society, it must be held to high ethical standards and yield trusted, verifiable results.

Nuix adheres to three key technology design principles when developing AI that are congruent with Nuix's purpose of building technology that is a Force for Good. These help to strike the delicate balance between maintaining the highest ethical AI standards in our solutions and optimizing AI's immense power for our customers. We advocate that these three principles must be at the core of the design and application of all AI, to ensure that community and societal trust is placed at the core of all current and future advances. These three principles are:

Explainability – To achieve and sustain confidence in AI's output, the 'black box' must be opened to provide humans with clear line of sight to the training data used to build the models, the biases that might exist, and easy-to-understand reasons behind the AI's predictions. AI should be used to inform and accelerate human decision-making, rather than purport to replace it by delivering definitive 'answers' and Nuix's AI points human decision makers to the places they are most likely to find the undeniable facts that will inform their actions.

Accessibility – The transformative power of any technology lies in how usable and customizable it can become. Nuix's AI ensures that the human stays in full control, and sets the laws of what the machine can, and cannot see and do on behalf of humans, without being beholden to inaccessible algorithms or code. AI that empowers – vs replaces – the human is central to this principle.

Specificity – The concept of specificity enables the pragmatic and efficient application of AI that drives measurable, real-world results. More generalist models are more difficult to explain and less transparent, also making them less repeatable. The key is to enable the users to drive the process by imparting their knowledge and expertise into the models they create, centered on one of a wide range of targeted use cases. The result is dramatically improved accuracy, and greater credibility in the domain-specific outcomes.

The implications of these principles, and how Nuix meets these requirements, are examined later in this paper.

WHY AI REGULATION NEEDS TO BE "LONG AND CAREFULLY EXAMINED"

While the cost of developing and training AI has plummeted in just a few short years, the computing and human resources needed remain significant. Significant AI advancements, therefore, have been concentrated within private enterprises. In fact, academia produces just one tenth of AI breakthroughs[2], meaning that much of the application and development of AI is likely being guided far more by profit motive than by ethical concerns.

Proof that the technology is evolving more quickly than society's understanding of its benefits and risks is the fact that even AI's leading innovators and advocates are calling for a more measured and cautious approach. Yet, while groups of business leaders publicly and collectively propose a moratorium on AI development with alarming frequency[3], no one wants to blink first.

Government and regulators endeavor to catch up with the pace of change, and consider AI holistically, rather than tinkering around the edges. The European Union, for example, is close to enacting one of the first comprehensive laws governing AI. The AI Act[4] passed a major milestone in the European Parliament just days before this paper was published, and the Act may be enacted as early as the end of 2023. It seeks to "ensure a human-centered and ethical development" of AI. Core to that proposed legislation is an "ability to understand what models were used to achieve the results, how those models were built (training data) and whether or not the human(s) in the loop are able to be the final arbiters of the outcomes."

As the EU led the way on data privacy, with the introduction of the GDPR[5], they appear to be leading once again on AI regulation. Taking the past as prologue, the EU's AI Act is a likely predictor of how AI regulation will unfold globally. Indeed, it is already part of some organizations' future planning. When a Nuix customer – a law enforcement organization in Europe – inquired about our AI, Nuix was pleased to offer detailed written assurances that our AI models and capabilities already adhere to the requirements.

In a more targeted approach to mitigating the potential risks of AI, a federal judge in Texas recently implemented a new rule that requires attorneys appearing before the court to submit a certificate that affirms that either no part of their filing was generated by generative AI (ChatGPT, Harvey.AI, Google Bard), or that any content that was assisted by AI has been thoroughly vetted by a human for accuracy.[6]

In short, realizing the vast potential of AI will take both serious thinking and thoughtful, coordinated actions. As two Australian AI academics wrote, "It's clear we're at an inflection point: we need to think seriously and urgently about the downsides and risks the increasing application of AI is revealing[7]."

"The saddest aspect of life right now is that science gathers knowledge faster than society gathers wisdom."

Isaac Asimov

THE NUIX VIEW

At Nuix we've been working with and developing AI for years. We understand AI has the power to transform our customers, and even to reshape entire economies. However, we also recognize that humanity may only have one chance to get the transition to an AI-enabled future right. Intelligent regulation, education, discipline, and honest debate are critically important to ensuring that AI is built and applied thoughtfully, ethically, and transparently. In short, Ethical AI is synonymous with Trusted AI.

As Nuix CTO, Stephen Stewart put it: "The reality is that leveraging AI/ML across the incredible diversity of data challenges out there is difficult. A one-size-fits-all solution for unstructured data doesn't really exist. Nuix's AI is an example of applying cutting edge concepts and technological advances in AI to meet very specific customer needs with incredible efficiency and defensible insights."

Our approach to developing and deploying trusted AI is governed by three technology design principles: **Explainability**, **Accessibility**, and **Specificity**. These principles keep us anchored in our purpose of being a force for good by finding truth in digital data. They ensure we create collaborative intelligence that increases human efficiency and effectiveness, enabling people to focus on the high-value work that requires human judgement, nuanced critical thinking, and ethical awareness.

These principles position Nuix as a responsible, reputable and ethical provider of AI. They also serve as thoughtful guardrails to guide our ongoing development efforts, and ensure we remain aligned with customer requirements and evolving global ethical AI standards.

1. EXPLAINABILITY

As global leaders in digital intelligence software, our customers rely on us to process and sift through, and interpret vast amounts of data to find factual insights – i.e. truth. To establish and maintain that trust in AI, our customers need visibility into the underlying data used to build our AI models, as well as the explanations behind the results produced by those models. This is increasingly known in the industry as 'explainability' and our AI technology was purpose-built from the ground up to provide it.

Another important element of explainability is the fact that our AI does not pretend to make decisions for humans. Rather, it delivers sifted, relevant, and prioritized data that supports more effective and efficient human decision making[8] by our customers. The key here is that the human has direct insights into each step of the process, and maintains full control as the final arbiter. Our AI offers four major analytic layers about textual data:

1. Record-type: e.g. tax document, healthcare form, transcript, patent.

2. Content Topic: e.g. law, environment, finance, politics.

3. Extractions: e.g. PII, PHI, named entities.

4. Priority: e.g. relevance, importance, or risk level of any of the above, based on customer-driven rules.

The Nuix solution points human decision makers to the places they are most likely to find the undeniable facts that will inform their actions and accelerate the accuracy and efficiency of their decision making. Every step in the process is recorded, repeatable, explainable, and defensible.

2. ACCESSIBILITY

Until recently, AI has been the preserve of a small number of highly skilled machine learning experts, software programmers and data scientists. These circumstances precluded domain subject matter experts (SMEs) out of the model building value chain and left them overly reliant on technical teams who often lacked the domain knowledge to work on those models.

To maximize the value from investment in AI, and also to meet the ethical imperative to spread the opportunities and benefits that AI offers, there is an obligation to empower non-technical individuals with access to those capabilities, in an intuitive, easy-to-use way.

Nuix is contributing to a user experience revolution, with an intuitive interface designed to enable domain experts to build, optimize, and validate highly accurate language AI models quickly – a simple point and click interface without a single line of code. Beyond the workflow and productivity advancements this offers, this level of accessibility deepens confidence that the AI's results are vetted and well-aligned with the subject matter for which they were designed.

Well-designed low- or no-code interfaces and integrated workflows help to democratize the power of AI, allowing a broader array of human workers to directly engage with AI to make needed changes, test and optimize models, correct errors, and remove bias. Allowing the user of AI – not the coder – to always remain in ultimate control.

3. SPECIFICITY

In the context of real-world applications, bigger doesn't always mean better. While ChatGPT is supported by GPT4 (the largest LLM ever built[9]), its immensity, and its attempt to be all things to all people, undermines much of its practical use in real-world business environments. Clearly it would be impractical, if not impossible, to host ChatGPT in your secure environment. And uploading sensitive data to use it 'as-is' on the web creates a material exposure risk. For this reason alone, many companies and organizations have prohibited employees from accessing ChatGPT and other large generative AI provider sites. These massive language models are not well-attuned to support narrow and deep subjects relevant to your business or subject domain.

By contrast, Nuix's AI is designed to be deployed in customers' secure environments, and run on affordable CPU-based systems. Furthermore, combining the advantages of explainability and accessibility described above, customers can quickly adjust templated models, or indeed build new models, to fit their subjective requirements and targeted use cases. Once again, the humans are in control, and the AI works to support their specific needs.

> "A computer would deserve to be called intelligent if it could deceive a human into believing that it was human."

Alan Turing

AI THAT PASSES THE TEST

All three of these technology design principles contributed to our prize-winning participation in the US Navy's CUI Tool Automation Challenge, announced in December 2022[10].

The goal of this challenge was to support appropriate, effective, and vital sharing of sensitive but unclassified information – Controlled Unclassified Information (CUI) – between departments, partners, and allies. The specific tasks to be managed included:

- Ingesting millions of US Government and Department of Defense documents in a range of file formats.

- Analyze the document contents to identify CUI types within each file.

- Quantify a confidence level in alignment with each result.

- Mark the banner and footer of each document to maintain the chain of trust as various humans interact with files.

The success of our solution was built on the foundations that our three principles provided. The **explainability** provided by a simple point-and-click interface into how the newly built and existing AI models were created, as well as easily understood reasons behind our results. The **accessibility** afforded by a no-code interface, which empowered the user to build, train, enhance, or adjust models and relative weightings to optimize and validate each model's performance. And the **specificity** of each model, carefully tuned for a specific kind of CUI, and combined by domain experts to address unique real-world data challenges.

AI AND YOUR ORGANIZATION

The ability to use advanced technologies quickly and effectively has long been identified by economists as more crucial to growth than the actual nature of the technology. Robert Solow's famous line is very apt: "You can see the computer age everywhere but in the productivity statistics".

At Nuix, we've built a global reputation making our customers' data highly searchable. Now, we are taking things to the next level with a unified, AI-infused platform called Nuix Neo. Nuix Neo combines the world's most powerful data processing engine with advanced AI to help solve our customers' complex and growing data challenges. Through this integrated innovation strategy, Neo customers can leverage the value of AI, without needing in-house AI expertise and without buying or building multiple point AI solutions.

SUMMARY

Delivering AI solutions is a task that shouldn't be taken lightly. In a world increasingly under siege by sweet-talking, hallucinating Chatbots and beset by misinformation, disinformation, and deep fakes, the stakes are higher than ever. But the reality is, the problems of today and tomorrow are simply not solvable with yesterday's techniques and technologies. Like it or not, we believe that mastering AI is part of our critical path to a prosperous and sustainable future.

The key will be a mindful and disciplined approach to AI, founded on Explainability, Accessibility, and Specificity. The ultimate goal will be to deliver trusted outcomes that are credible, verifiable, and defensible for our customers, and society at large.

WANT TO FIND OUT MORE?

The trends covered in this whitepaper are still evolving – as are the real-world applications our customers have for Nuix's AI capabilities. If you'd like to discuss how those capabilities can work in your organization, reach out to us using the channels outlined below. One of our AI experts will be in touch to help you understand how Nuix Neo and its embedded AI can work for your organization.

REFERENCES

1 Krishna, Campana and Chandrasekaran, 'Thriving in an AI World: Unlocking the value of AI across seven key industries,' KPMG, 2021, https://advisory.kpmg.us/articles/2021/thriving-in-an-ai-world.html

2 Stanford University, 'Chapter 1.2., Research and Development', Fig 1.2.2, Artificial Intelligence Index Report 2023', 2023, https://aiindex.stanford.edu/report/#individual-chapters

3 See https://www.science.org/content/article/alarmed-tech-leaders-call-ai-research-pause#:~:text=Attracting%20 signatures%20from%20the%20likes,potential%20risks%20of%20 the%20technology

4 See https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence

5 See https://gdpr.eu/what-is-gdpr/

6 See https://www.mitchellwilliamslaw.com/taming-the-ai-beast-judges-set-rules-to-control-use-of-generative-ai-in-their-courts

7 Sonenberg and Walsh, 'Artificial intelligence is now part of our everyday lives and its growing power is a double edged sword', The Conversation, November 2021, https://theconversation.com/ artificial-intelligence-is-now-part-of-our-everyday-lives-and-its-growing-power-is-a-double-edged-sword-169449

8 As we write, Nuix is in discussion with one of Europe's largest law enforcement agencies who are looking at how they can capture the benefits of AI but with a core focus on the transparency of outcomes

9 See https://neuroflash.com/blog/gpt-4-parameters-rumors-and-forecasts/#:~:text=It%20is%20also%20possible%20that,trillion%20 or%2010%20trillion%20parameters.

10 'Nuix and Serco NA partner to score prizewinning results in US Navy AI Automation Challenge', Nuix, https://www.nuix.com/ news/nuix-and-serco-na-partner-score-prizewinning-results-us-navy-ai-automation-challenge

nuix

Nuix (www.nuix.com, ASX:NXL) creates innovative software that empowers organizations to simply and quickly find the truth from any data in a digital world. We are a passionate and talented team, delighting our customers with software that transforms data into actionable intelligence and helps them overcome the challenges of litigation, investigation, governance, risk and compliance.

APAC
Australia: +61 2 8320 9444

EMEA
UK: +44 203 934 1600

NORTH AMERICA
USA: +1 877 470 6849