techellect™
AI and IQ working together

Safe & Secure Use of AI Tools - Vol 2

From the Desk of Dr. James Norrie

Mastering Emerging Technology: Principles for Positive Impact in Personal and Professional Realms

So, let's explore three principles, including examples, across each of these two specific use cases – personal and professional – to see if we can establish some fundamental early guidance you can rely on to really help you master this emerging technology for good while avoiding the bad!

Principle #1: Be AI Safe & Secure

As with any technology that engenders some end-user risk, caution in the use of AI is advised so that you remain personally and professionally secure and safe while using these platforms. As we explain in our AI training modules, be aware that every time you query an AI engine, they immediately possess and control all of that new information. It is theirs. In fact, the underlying technology platforms – large language models or LLM's – are built on the premise that with use over time, these queries and the value and reliability of the response – are part of how AI platforms learn. So never share anything that is confidential, personal, sensitive or inappropriate with any open, public AI system.

Personal Example: while it may seem secure and safe to share personal information with an AI platform, once what you share enters the LLM (large language model), it can never be recovered, updated or expunged. Therefore, be careful with any query that contains any personal information. If you have children, it is equally important to help them also understand these risks that they may be entirely unaware of and take for granted unless parents and teachers educate them accordingly. If you need more help on this, check out our free AI training content for both adults and children to learn more.

Professional Example: employees may be tempted to use AI to help them with their professional tasks. However, they must be instructed and educated on the risks of doing so when any information required to complete the task may disclose restricted, private, or confidential information or anything that might involve trade secrets or intellectual property for instance. Instead of banning the use of these tools and driving their use to employees' homes or personal devices – which studies suggest is already happening – consider developing and implementing corporate guidance on the approved uses of AI including guidance on how to safely deploy and use these tools accordingly in a workplace setting that protects confidential and private information from leakage. If you are responsible for this aspect of your company's operations and/or a business owner, check out our free starter AI policy pack for some help.

techellect™
AI and IQ working together

Safe & Secure Use of AI Tools - Vol 2

From the Desk of Dr. James Norrie

Principle #2: Demonstrate Provenance Integrity & Accountability

Steven Schwartz, an experienced attorney, admitted to using ChatGPT to prepare a brief in his client's personal injury case against Avianca Airlines (NY Times, Reuters, et. al 2023). For details on this interesting early case, just search for it online. It's easy to find this sensational early example!

In addition to not having demonstrated proper legal responsibility for his work – a breach of legal ethics and integrity – he failed to internalize an important point about AI tools. Unless you have knowledge of or control over the data provenance of what was used to populate the LLM model underlying the GAI tool you are using, do not blindly rely on it being accurate. Just as in the case of information being online not all being true, AI too can be improperly trained to even directed to not reliably produce only correct answers. In fact, as we speak, cybercrime gangs are already using AI against us in social engineering to advance their criminal agenda by programming it to lure and lull you into to do something you shouldn't. Right now. So, the skeptical attorney in question, added insult to his own injury by using the same GAI tool that prepared to fake brief to ask if its contents were real. How well do you think that went? Of course we see in hindsight that he should have validated this AI output himself. That would be demonstrating professional accountability for your work. Relying exclusively on AI is not. As can be seen, while the information that the LLM returned seemed reliable and reasonable, every single court case that was quoted turned out to have been a fictitious amalgam returned by ChatGPT to satisfy his query rather than to produce valid legal research creating a dangerous outcome.

While OpenAI – the so-called "guardians of ChatGPT" - quickly moved to detect what happened and how within its platform to ensure a similar thing did not happen again, I am quick to note that with AI being "self-programmed" – just like a human – it and its creators may learn best from its mistakes that its successes. Just as we often do. So, do you want to be that guy or gal who used ChatGPT to take a short cut only to learn that your lack of accountability for disclosing that and having the integrity to properly conduct yourself in the use of this platform costs you your personal reputation and/or even potentially your job or career? If you use these tools, exercise your own personal judgment instead of relying exclusively on the GAI tool to ensure that what you stand behind is what you personally verified and validated and, where appropriate, provide disclosure of your use of this tool as one of many reference tools or resources you used in preparing the materials especially if you are relying on those in a professional or work setting.

Personal Example: Currently, there is a raging debate among teachers and professors about what to do with kids who "cheat" by using GAI to prepare and submit graded work. In fact, some students have taken to calling it "CheatGPT" instead of ChatGPT as a result. This returns us to a basic and fundamental human question: is your own integrity worth losing because you get caught cheating using these tools? Does your integrity even allow you to pass someone, or something else's work as your own? If so, welcome to the new world of GAI where you will be in cheaters' heaven. However, if like most people, you value your integrity, then find ways to yourself accountable to a higher standard of being aware of where the information provided is coming from, ensuring its reliable and real, and not over-relying on the judgement the tech instead of your intellect!

Professional Example: For business owners and executives, the risk in this area is substantially more devastating than the personal example provided. Any company is wholly responsible for its conduct and the conduct of any of its employees or agents. This means that if an employee were to access and use chatGPT, for example, in the preparation of anything that you were publishing or promoting and it turned out that the information was unreliable or even completely false, you could end up being held legally liable for that outcome. On the other hand, GAI tools can be very helpful in a number of areas of your business (HR, customer service, IT help desks, etc.) so you do not want to simply ban them. Instead, explore the high value use cases inside your own company, explore how you can put safeguards in place in collaboration with your employees who use these tools in their work, and ensure that the entire company understands the risks of misrepresenting information from ChatGPT and other tools.

techellect™
AI and IQ working together

Safe & Secure Use of AI Tools - Vol 2

From the Desk of Dr. James Norrie

Principle #3: Understand Information Warfare & Disinformation

For centuries, nations and societies have understood the power of information. And also the power of disinformation. Today, with the easy access that the internet provides to each other, information warfare is a deliberate and continuing battle being waged among superpowers and wannabe's in order to get your attention and influence your thinking. Is there still any doubt that the Russians tried to interfere in our elections, and will always try and do so to their advantage? Or that the Chinese – already the 2nd most powerful nation in the world economically and technically – intend to replace the US as the world's leading super power? Now imagine that you had a new tool in your information warfare toolkit – and that AI could be used to engage in persuasive conversation and rhetoric that was actually entirely designed to sway public opinion on something for the benefit of another nation? Well, don't just take our expert word for that, research what's really happening (CISA, 2023 for example) and draw your own conclusions. The more compelling a conversation is – especially online – the more persuasive it could become among those not aware enough of the tactic to avoid being influenced to criminal, terroristic or similar activity. Or less dramatically but still important, to believing and propagating dark conspiracy theories that are ultimately intended to undermine democratic institutions. Or how about simply exploitation of your information for criminal gain?

Personal Example: As a citizen you take your responsibility to vote seriously and are diligent in exercising your civic responsibility. You consume much of your opinion influencers online and are always looking for new sources and angles to help you think through the politics of the day. You look to a new AI-enabled search engine which you have found really useful in locating lots of new information. However, of late, your summary of politics has become more and more frustrating as detect more division and strife than actual results. You decide that voting just isn't worth it this time around because it won't make any difference anyway, right?

Meanwhile, the back story is that other nations have actively planned much of what you consumed to deliberate appeal to your emerging frustration in the hopes you won't vote while motivating others by doing the opposite to them and ensuring they get angry enough to vote. Their goal? To elect a US government through indirect online influence that could be to their advantage and would be more malleable globally.

Professional Example: In a business setting, similar tactics are used but they are typically more subtly deployed, often entirely without your knowledge. For example, your company operates globally but is based in the US and must adhere to guidelines on its conduct such as the Foreign Corrupt Practices Act (FCPA). You use AI to search for this and to gain guidance on an issue of concern to you where internal knowledge was not present about. As you type the query, the tool is actually drawing on false information planted within the LLM by other parties and returns a result that directs you to what seems like a legitimate source of advice on this delicate issue. You engage, and in the process of doing so, begin to disclose information about your bid on the project in question that is of value to foreign competitors without even beginning to imagine that this was all a ruse designed to draw you in and not to help you out.

Conclusion

While many of the examples here are given as a warning of the risks, it is important to conclude with an important corollary thought. AI also has many potentially life-changing and beneficial social impacts yet to be discovered and deployed. Good too shall come of this amazing technology advance.

But in the early going, what is paramount for us to all do is remain human centric in our pursuit of these benefits, seeking to ensure that there is transparency, fairness and disclosure to ensure that AI companies and their platforms are held to account for not just producing profits but also for being socially beneficial, because previous assurances of self-regulation have failed to address the essential greed that undermines any capitalist system and economy.

Businesses and organizations are going to have to adapt quickly to the fact that AI is here and never going away. An initial risk assessment of anything new often causes large enterprises especially to pull back because they cannot see how to embrace this new tool without it being disruptive and risky. However, the smartest organizations will be those that quickly move to figure out how to buy or build their own AI applications and to harness these new tools for the benefit of their investors, employees and customers alike.

Governments will need to embrace new social outlooks, policies and practices that can help us harness AI for good and to use legislation and regulation to stamp out the bad where that is possible to achieve. They may wish to consider new ways of approaching international co-operation or treaties to deal with the unique complexities of AI platforms that are globally owned and deployed instead of defaulting only to national interests. The daunting task of harnessing AI for human good affects every human on the planet, not just those who life in a specific country or region. We will need global solutions to potentially global issues. And cooperation will be essential to achieve positive outcomes.

Following that, our underlying legal system – by definition mostly backward looking – will need to move swiftly to address early legal cases with a thoughtfulness about their human and societal implications and to judge them within a fast moving and innovative realm that is going to be challenging to regulate and control. We must ensure that our legal tools keep pace with technology innovations, apparently including how to use those tools in the practice of law appropriately versus inappropriately.

Educators need to ensure they are doing their jobs around information literacy, digital citizenship and ensuring that GAI is introduced early into the curriculum so that humans master its use to help abate fear, lower anxiety and improve the bytes and brain collaboration that AI invites. This is not a time to fear change, but to embrace it. Education will no longer be about the accumulation of knowledge but about the power of information in context used for decision-making. This will also challenge all kinds of existing jobs, occupations and career paths while opening up brand new opportunities for new kinds of work and social advancement that the earliest roots of education helped to create and continue to create today. Educators need to embrace their role in helping humans master AI and not fear it as a challenge they cannot master themselves, much less with their students.

And the very idea of the ideal human experience is going to change in as yet to be seen ways that we cannot entirely forecast or foresee yet today. As with all important technology advances, there will be many bumps along this early road to more prosperity, and some important risks to reign in quickly. But ultimately, AI should be a compliment to human intellect not a replacement for it. So now the hard work begins in each and every one of us: what are we each going to do to individually embrace this emerging AI reality, to humanize it, and to master it over generations to come for the betterment of all? For that will always be the best path forward for humanity as a whole.