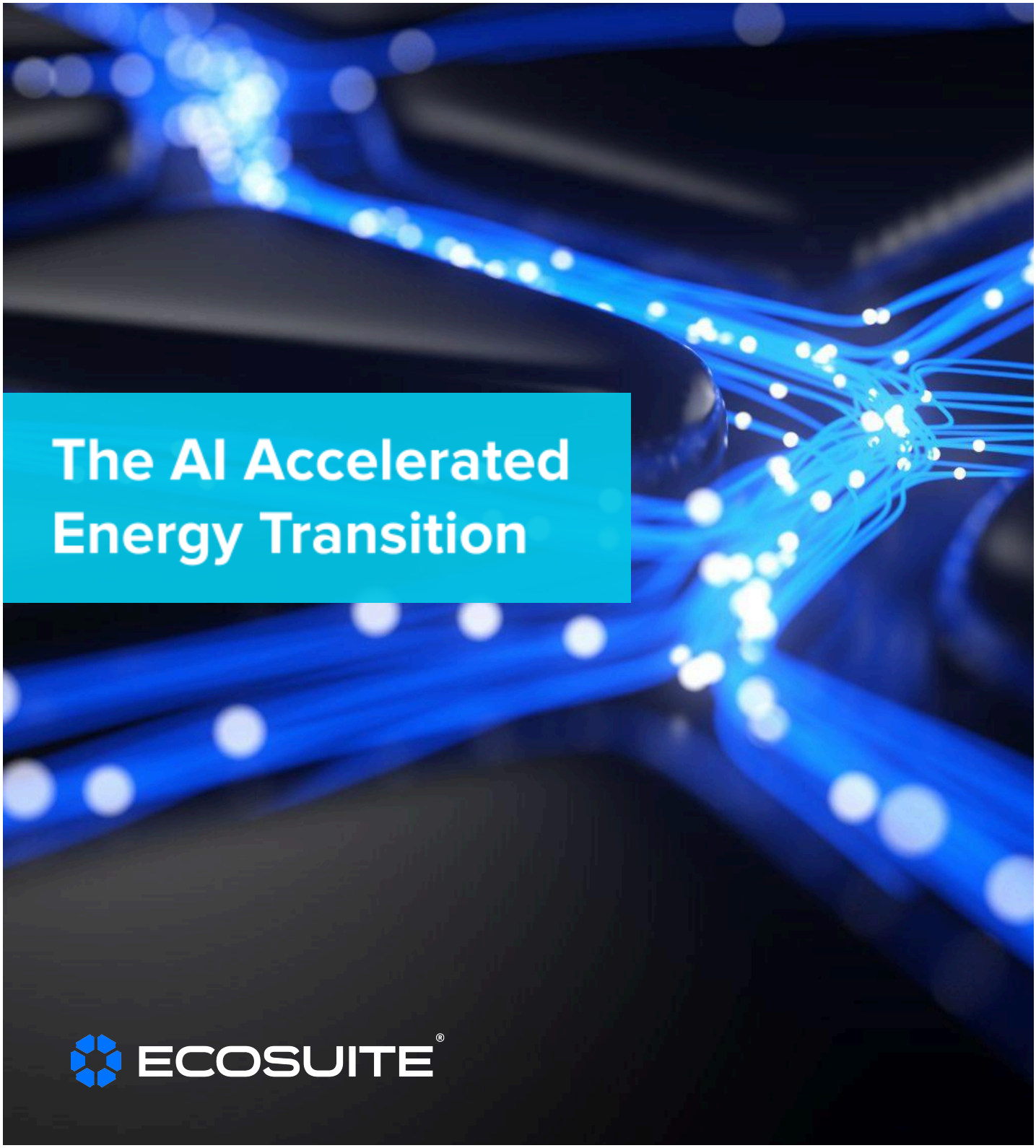


Author's Note: this white paper assumes the reader has a base knowledge with respect to Artificial Intelligence (AI) *technology, if you are new to the ideas of Neural Networks and AI, Wikipedia is a good resource and can serve as an effective primer: https://en.wikipedia.org/wiki/Artificial_neural_network*

Philip Court and John Gorman,
with contributions and support from Jack Bertuzzi, Isabel Wilson and Joel Santisteban

INTRODUCTION

Machine Learning (ML) and Artificial Intelligence (AI) are making daily headlines, with many futurists and technology writers claiming that they will change entire industries and economies. Based on demonstrations of products like ChatGPT, the capabilities are indeed remarkable. While machine learning, based in neural networks, is not a new technology, the application of these methods scaled up with the latest computing infrastructure has allowed Large Language Models (LLMs) to deliver something tangibly new.

AI's many promises are now cultivating dreams of a utopia where friendly digital beings are knowledgeable and omnipresent; on the flip side, these same digital beings can stoke fears of out-of-control, potentially dangerous AI systems. Most industries have of course taken a positive view and are looking to leverage the capabilities that AI developers and providers are busy selling to the world. Renewable energy and Distributed Energy Resource (DER) deployment will not escape this trend as the drive for optimization and efficiency continues.

The Ecosuite team have been waiting and preparing for this moment since the early 90s. It's been a slow progression in some respects, mostly due to constraints of computational infrastructure, but the ideas and latent expectations of what might be possible have now arrived. Consequently we have not been waiting around.

So far we have two AI irons in the fire and in use by a small carefully selected group of collaborators. These two separate technologies are:

1. **SolarQuant**, a tooling that allows ML experts to easily train and deploy ML algorithms to the grid edge.
2. **STEVIE**, an Ecosuite feature allowing admin users to instantiate LLMs tuned to assist specific roles.

Both of these are explained in further detail in the sections below. For now, suffice to say, we are stepping boldly into the utopian future as we pursue the clear opportunities to accelerate DER deployment, the energy transition, and of course the overall effectiveness of DER teams everywhere.

SolarQuant

Makes it easy for machine learning experts to leverage large DER datasets, using their traditional tools of choice to train and deploy ML solutions to the grid edge with zero friction.

One of the most interesting experiments with machine learning is to “set it loose” on a large dataset, in order to find valuable patterns that a human could not necessarily perceive. Ecosuite users are building these datasets and they are growing by the day. These datasets are a result of maintaining multi-vendor footprint DER portfolios with acute attention to detail about the asset health within these portfolios, this health is fundamentally gauged by examining the data each DER radiates. Managing DER portfolios does not fit under a “one-size-fits-all” approach – reality means managing different makes and models of equipment from a range of verticals.

In general this translates to having to deal with a range of solar inverters, energy meters, pyranometers, EVSE, BESS, BMS and more. These devices are manufactured by different companies over different eras, using various communication protocols, network topologies and software interfaces – and yet, they all need to be managed coherently to extract full performance from a portfolio. Building upon open-source software frameworks, and a Bring Your Own Device (BYOD) model, the Data Acquisition Systems (DAS) prescribed by Ecosuite are able to easily manage this varied and disparate asset base, using interoperability and flexibility to provide normalized, high resolution time-series data securely from the field.

These DAS devices serve as a solid, repeatable “cookie cutter” building block for the distributed energy solutions that industry requires. The DAS not only acquire data from the field, but they can also receive remote instructions to execute. These instructions can come from centralized decisions in the cloud (e.g. from Utility specific OpenADR integrations) or more directly via distributed logic at the grid edge. Like a Swiss Army knife, this modular and open software supports all DER assets and is deployed into locations with many variations in networking topology. From here the DAS delivers a clean, continuous, high-resolution dataset to be leveraged within Ecosuite, employing bidirectional SSL certificates for best practice security and nonrepudiation. Additionally, when new technology and devices emerge on the market, these DAS can accommodate and support them quickly (it takes approximately 5 days for a new device to be added to the supported list of plugins). **Known within the Ecosuite team as “silo-busting”, this process of rationalizing asset information helps to lower soft costs and deliver unique datasets primed for new value creation via machine learning.**

Managing assets now becomes more efficient because the asset management software, which is a small part of Ecosuite, now exhibits a highly uniform and organized dataset. This growing dataset is the perfect hotbed for machine learning to unlock further value from projects and portfolios, offering the opportunity to deliver increased performance and new services to customers.

The software framework we have created to realize this is called SolarQuant. SolarQuant (which works with and is part of the open-source solarnetwork.net software platform) delivers a general, repeatable and configurable ML framework focused on streamlining and accelerating ML researchers’ work and also allowing easy deployment of their results to the grid edge via ML solution repositories on GitHub. Importantly, this

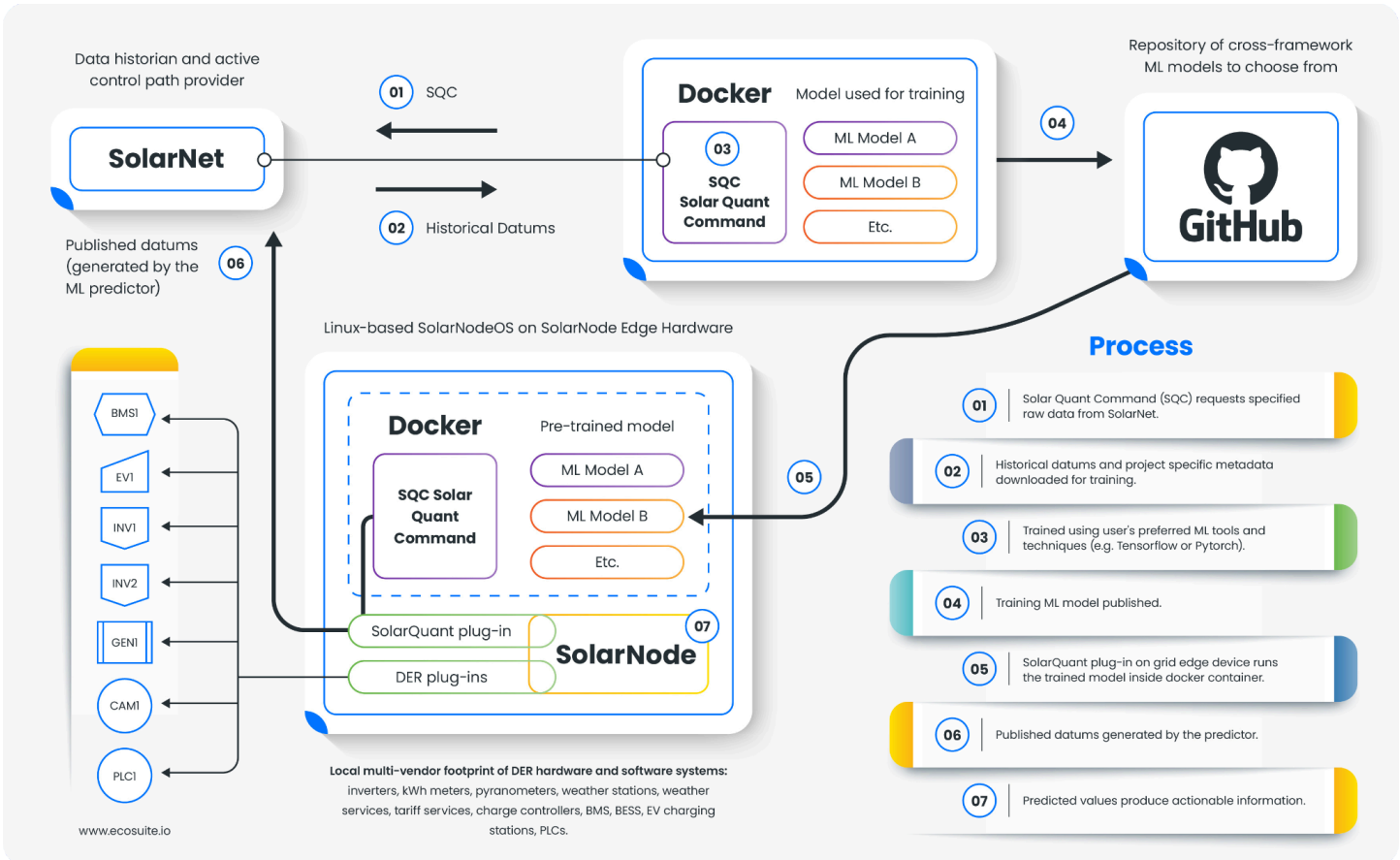
framework is technology agnostic; researchers can leverage whatever tools and approaches they like or prefer to use for the problem at hand.

The challenge of building SolarQuant came down to: **how do we support all these efforts with a machine learning management engine that is agnostic to computer language or framework?** We noticed that a lot of the most compelling ML work was already being done using open-source frameworks like TensorFlow, PyTorch, scikit-learn, emergent and others. We also knew that advances in ML applications within knowledge domains often came through trial and error, lots of data crunching, reviewing results and tweaking the methods – the same iterative process we knew worked well for software development.

We realized that if we could create software infrastructure that let data scientists use the tools that they were familiar with, but standardize access to data and the deployment of the resulting ML algorithms in flexible and automated ways that this would be a game changer. With this, empirical experimentation could proceed, leveraging improvements and testing in parallel across many different datasets, and scheduling of training and prediction could proceed smoothly via automation. By open sourcing this software framework, we could allow different teams from around the world to experiment and share results on the best ways to solve the problems DER asset owners need and want solved.

In summary, this is what SolarQuant is: A framework for easy access to data, providing a zero constraint environment for ML, including model design and training and then an easy mechanism for publishing and deploying the resulting trained models.

Here is a process flow diagram of SolarQuant and its components, including Solar Quant Command (SQC):



Further details on SolarQuant and SolarNetwork are available here:

- https://ecosuite.github.io/solarquant/
- https://solarnetwork.net

Applications for machine learning on DER datasets are varied, but current goals include:

- Fault detection (identifying the specific causes of equipment faults in real time)
- Preemptive fault detection (predicting future equipment faults via incipient fault detection)
- Predicting and validating weather adjusted performance based on site-specific system history

SolarQuant makes use of containerized application development and deployment – specifically Docker – to package, deploy and replicate ML configurations easily. What this allows, is to not only scale ML compute in the cloud, but also to deploy it to the grid edge as an alternative and more organic form of scaling the overall solution's compute needs.

The architecture is also designed for **iteration** and **inclusion**. Not only will different researchers be able to download the software, use their own codebase to develop methods and cascading process flows, but we believe over time, a GitHub repository of models – published by ML researchers, for all to use – will accumulate some of the best trained models within the DER industry. This collection of models will be tested, validated and ranked by researchers and DER owner operators alike, for the benefit of all. We look forward to collaborating with other open source projects such as Solar Forecast Arbiter, pvlib, and SunSpec OrangeButton to better distribute and validate the learnings that are made through experimentation and iteration. One of our immediate goals with SolarQuant is to reduce soft costs in the Operations and Maintenance (O&M) space, and hence let C&I DER and solar deployments maximize their potential.

STEVIE

Empowers team members to instantiate and use digital assistants with specified AI personalities tailored to assisting certain tasks and roles within their company (achieved by leveraging LLMs enhanced with carefully selected subsets of DER project data, ingested via programmatically defined templates or "AI Personalities").

Our Seriously Talented Ever Vigilant Industry Expert (aka STEVIE) is another way we are empowering people with AI. Teams use STEVIE to accelerate tasks that need to be done in their normal day to day business, i.e. developing and deploying DER projects! With generative AIs using large language models (LLMs), ChatGPT being a well known example, coherent human-readable expressions of knowledge can be generated in an almost immediate and seemingly flawless fashion. Although, due to the nature of how LLMs work, we know that they can sometimes provide answers that are not true. For us, this is problematic, because we are always looking to hold the "source of truth" for DER projects as an immutable fact at any moment in time (e.g. milestones, design details, construction details, costings, asset performance or return on investment). This information is critical not just for Ecosuite's DER developers and their internal teams, but even more so for their external investors and partners. The expectations and needs for a DER owner/operator are much more specific than that of a chatbot user (who might still have entertainment as their primary goal) – the breadth of knowledge demonstrated by a base LLM is useful, but more important for Ecosuite users is the domain-specific and project-specific knowledge, i.e. everything there is to know about DER assets in a portfolio, their

construction, their provenance and their performance. With a variety of LLM vendors and models available to use, software developers have an ability to pick and choose between these base LLMs and also to augment them with domain-specific knowledge via the use of tokens, this lets us to some degree tune a model to provide more exact and relevant domain specific responses.

Via programmatically defined prompt engineering and the tokenization of data for DER projects within a company's portfolios, we can augment the base LLM of choice with portfolio-specific knowledge, shaping and improving these LLMs to better serve our users' collective needs. Hence we now have generative AIs that can rapidly inform engineers, project managers, asset managers or contractors in their day to day jobs. Ultimately this saves time and guides teams to make better decisions.

By wrapping a range of LLMs in a software framework that lets users instantiate new AI personalities, we put users in a very powerful position. By embracing the fact that they can now define and shape new AI personalities over time (by refining datasets via tokenization and the Ecosuite's embedded programmatic prompt engineering), we are empowering Ecosuite users to not only use AI, but to provide feedback for the purpose of ongoing iterative improvements. The sky's the limit, and we intend to extract every benefit possible from these new capabilities. Initially these benefits are being reaped by a few selected beta users (additional Ecosuite users will be granted access at a future date).

This technology will continue to evolve and improve, but for the time being these are some of the areas and capabilities that this technology is offering and assisting teams with:

- Offering resolutions to particular situations (e.g. Asset Management)
- Offering opinions on complex business decisions (e.g. Investment Committee Approvals)
- Answering questions of any form about any project, portfolio or any aspect of the business
- Answering general knowledge questions (e.g. about the industry or more broadly about society)

With the right feedback loops, increasingly large datasets, and ever-larger processing power at our disposal, AI will continue to exceed and more likely *dramatically exceed* the vast majority of human expectations. Looking at it from a distance, that really is our hope. Our hope is to amplify teams using Ecosuite and, by extension, the wider industry and society's effectiveness in accelerating the energy transition. As long as we continue to shape this technology so that it enhances team culture and empowers teams to bring more speed, more innovation, and more time for human creativity and thoughtfulness, then our collective decision-making will be better for it. Ultimately, our collective results will be amplified and accelerated.

OPPORTUNITIES AND RISKS

We are stepping boldly into the future. In these early stages we are already seeing satisfying results and clear opportunities for future gains, but what about potential risks? Have we considered this? Are we on the path to creating an AI that leads to a dystopian nightmare?

These are not stupid questions to ask. We must ask questions of this nature and defend against negative possibilities, especially because AI can indeed be put to nefarious uses. But rather than worrying about AI with explicitly malicious intent, here we are more focused on the possibility of a benign AI going rogue and going rogue in such a way that it is unexpected and hence will not likely have been defended against by the system designers. This is not out of the question, as there are many examples of this exact problem already

happening[1]. In the case of a chatbot going rogue, this can be quickly mitigated and the consequences, although very real for those on the receiving end, are limited to some degree by the fact that it's just a chatbot. Hence, the more interesting problem is to consider the consequences when an AI is more than just a chatbot. What happens when the AI is connected, when the AI has interfaces with the physical world, more than just read only interfaces, but read/write interfaces, i.e. interfaces where the output from the AI can have direct impacts on the physical world (for example, changing the energy consumption of many DERs, hence impacting grid stability). An AI with these sorts of connections needs to be treated with a lot more respect and caution. This is of interest because we can foresee a future in the renewable energy industry in which AI is granted capabilities like those described above. If these AIs are designed to iteratively learn from their interactions with the world (i.e. retrain with the latest data on a regular basis), then we need to be even more careful, because without supervision and continuous verification testing, this is where unexpected or undesirable results could manifest. Either way, just like with cybersecurity, where there are many industry best practices, data validation being just one, and these best practices evolve with time, the AI industry will need to develop a similar set of best practices and this is already underway. Taking this analogy full circle, one of the known weak links with cybersecurity is the human element (i.e. a phone call or an email encouraging someone to do something that they really should not be doing – phishing is a prime example here), hence with AI security, that human element will also be part of the equation and something that needs to be considered. Looping this back to the chatbots that have already gone rogue, we see that conversational human language interfaces are not necessarily so limited or benign as one might initially think. A rogue AI can definitely impact the wider physical world via human influence in a chat room and hence needs to be considered a genuine threat in the bigger picture.

How does Ecosuite deal with this challenge? We have multiple strategies in place to address this:

1. Adhering to existing AI security guidelines (e.g. OWASP - Project AI Security and Privacy Guide)
2. Ecosuite alignment – our programmatically defined templates (personalities) have by design, aspects of "Alignment" built in. We tune these personalities to provide desired results just like the providers of our base LLM solutions also do.
3. STEVIE is currently only accessible to selected beta users, hence limiting the AI surface.
4. Users are trained and informed on the nature of STEVIE. This nature includes the fact that AI responses can be wrong.
5. Users of STEVIE are educated to interact as they would with any other member of their team, this means interacting politely and with respect, i.e. no abuse or inappropriate behavior as this can and will lead to negative and potentially unexpected results.
6. **None of Ecosuite's AI tools have direct control of physical infrastructure.** All physical system changes or actions that may be influenced by an AI are initiated by people trained in the nature of AI. In short, we always have a human in the loop where AI is involved. Any change to this policy would require an extreme level of data validation combined with other mitigation strategies which would come from a broad and deep risk analysis of the proposed AI driven automation (which we are not yet ready for in these early stages).

So, this means users will treat instances of STEVIE as if they are an intelligent and very knowledgeable new hire to their company (i.e. excited to have the new energy and talent on the team), while also being cautious of not being overly reliant on their input. The real difference being that this new hire has a lot of company-specific knowledge under their belt on day one (since they have already ingested all of the relevant portfolio information and associated data that their assigned personality prescribes). It's an interesting mix!

[1] https://www.theverge.com/2023/2/15/23599072/microsoft-ai-bing-personality-conversations-spy-employees-webcams

CONCLUSION

So now that users have instantiated a few of these AI personalities and deployed selected AI to the grid edge, what comes next? Now we have empowering **digital colleagues** who can confidently answer any question within the domain of their personality with a high degree of accuracy, and they themselves will be empowered via grid edge deployed AI generated data. Although just like our human colleagues, our digital colleagues can still be wrong on occasion, no matter how confident they may appear – we must never forget this fact. Hence, for decisions that really matter, double checks should be made. In general, our digital colleagues can now amplify our effectiveness and help unblock what would have historically been blocking or delaying questions. They accelerate our daily tasks to deliver faster and more regular customer satisfaction. Ultimately, this means delivering higher returns and better business results.

How far can we take this (and in this context 'we' means all of humanity)? There seem to be no obvious fundamental limits to how far this can go. The choices though, are ours to make (as a collective human society), but due to economic drivers, it's unlikely that this AI development and evolution will stop anytime soon. So how long until we have our first robot CEO? Likely not far away, but truly effective and useful CEOs bring many human elements (like a real human presence) that AI is unlikely to be able to emulate any time soon.

In a future white paper, we may go deeper on this subject (please reach out and let us know if you think we should). It could potentially touch on a number of topics including Universal Basic Income (UBI), our finite planet, limits to growth, upgrading economics, computational irreducibility, humans becoming multiplanetary, steady state economics, the near infinite universe and our growing but rudimentary understanding of what it means to be conscious and alive. There is no question that AI sheds new light on all of these topics, hence it definitely has something important to teach us.