



18 May 2023

Dear Co-Investor

SaltLight SNN Worldwide Flexible Fund (Class A1)

<i>Period</i>	<i>Return</i>
2020 (starting 11 November 2020)	+3.42%
2021	+18.09%
2022	-33.58%
*** 1Q	+11.69%

During the past quarter, our fund's Class A units increased by 11.69%. This growth was largely driven by the appreciation of our technology investments in the United States and China. We also managed to steer clear of potential losses from our long-term holding, Transaction Capital, a move which we will elaborate on later. Our "Resilience and Optionality" process played a crucial role in a series of fortunate actions that led us to completely sell down in early January.

This letter is going to focus predominantly on a topic that we've had our eye on for quite some time: Artificial Intelligence (AI). We've been writing about AI since 2021 and have been allocating capital to it since the inception of this fund.

New technologies tend to cause a ripple of excitement, and while these breakthroughs can bring immense benefits to society, they do not always translate into long-term financial gains for investors. Our aim with this letter is to blend our thoughts on AI, sift through the hype, and examine the 'investable' opportunities that this technology presents. Prepare yourself, as this letter is a bit lengthier and more technical than our usual correspondence. We believe, however, that it's important to provide you with a comprehensive understanding of our current approach to this emerging field.

Before we delve in, it's important to note that we are still in the process of learning. We are aware that attempting to predict the future is a bit like trying to catch a swarm of bees with a butterfly net. You might snag a few, but there's a good chance you'll get stung more often than not. And the one thing you can be absolutely certain of is that it's bound to be a buzzing chaos. We may not unfailingly nail every technical concept. And, of course, there's always the possibility of some new-fangled technology or business model peeping over the horizon that might deal with our stated problem in a slicker fashion.

Nonetheless, we remain somewhat stubborn in our efforts to put our thoughts down on paper. Please understand that our viewpoints are based on what we know today, leaving the door wide open for our perspectives to evolve.

This is a New Technology Epoch

The current breakthroughs in AI are reminiscent of those game-changing moments in innovation history that have profoundly shaped our world:

- The astonishing 1 minute and 45 second flight by Orville Wright in 1908 in Le Mans, France. This brief 'airborne' moment heralded the dawn of the aviation era ([video](#)).
- The introduction of the assembly line by Henry Ford in 1913, was a revolutionary concept that dramatically improved the cost structure of the automobile industry. It was this assembly line that brought down the cost of a Model T from \$825 in 1909 to just \$345 in 1916¹;
- Steve Jobs' unveiling of the iPhone in 2007, was a seminal moment that signalled the future of mobile technology and ignited a decade of innovation and new business creation built around mobile devices ([video](#))

Our approach to this new epoch is to apply our tried and tested investment framework while drawing upon our understanding of history. Previous technology epochs — such as railways, steel, mass production, information technology, and the internet — provide valuable lessons for discerning investable opportunities.

Our journey with AI began more than a decade ago. Over this period, we have persistently worked to broaden our knowledge and understanding. We have built AI models, completed advanced courses in deep learning, experimented with a multitude of pre-processing and processing software solutions, and sought advice from experts and companies in the field. Despite our relentless efforts, the rapid rate of AI development presents a formidable challenge to stay ahead of the curve.

Briefly, let's talk about **Transaction Capital (TCP)**. In 2022, we decided to gradually reduce our substantial investment in TCP. In our former fund, our initial entry into the company was at a price of R11 per share. Over the years, TCP has made a significant contribution to our overall performance. However, when management acquired WeBuyCars, the share price climbed to levels that we considered overly optimistic. This prompted us to reassess the future prospects of TCP through our "Resilience and Optionality" framework. The lofty valuation appeared to anticipate optimistic outcomes, while also presenting significant downside risk.

As we moved into the last quarter of 2022, we started to become increasingly concerned about several macroeconomic headwinds affecting TCP:

- *Fuel costs*: The sharp rise in fuel costs threatened the ability of taxi drivers to meet their payment obligations.
- *Floods*: Unusual flooding in Durban had a detrimental impact on Toyota vehicle production, which in turn hindered new vehicle sales.
- *Deteriorating book*: TCP's public securitisation reports indicated a noticeable degradation in the quality of their loan book. Further, management had

¹ Source: https://en.wikipedia.org/wiki/Ford_Model_T

communicated their intent to "term out" loans and alter their curing policy, both of which were worrisome signals.

In the early weeks of January, we decided to completely divest our position, with the hope of re-entering the market at a more favourable risk-reward level. However, we could not have foreseen the subsequent share price declines that occurred in March. With the share price returning to levels close to our original entry point from many years ago, we began accumulating a position once again.

We have tremendous respect for Transaction Capital's exceptional local management team. Our longstanding communications with them have demonstrated their entrepreneurial prowess time and again. We are confident that they will adjust to the changing circumstances and continue to thrive in the long term, just as they have in the past. While the journey ahead may be fraught with turbulence, we are thankful for the opportunity that the market has once again presented to us.

Now, back to AI...

The Breakthrough That We've Been Waiting For

In [May 2021](#), we shared our thoughts on AI with a particular focus on NVIDIA, expressing our belief that it was poised to become a significant enabler of AI technology. At the time, it felt like our insights were being carried away by the wind.

Our thesis hinged on the idea that NVIDIA's GPUs would form the fundamental computing hardware for neural networks. This hardware would not only develop AI tools and infrastructure but also democratise access to AI technology.

However, we hypothesised that a 'complexity bottleneck' might stymie the widespread adoption of AI. Deep-learning models, we thought, require expertise that only PhDs typically possess, and the true unexplored opportunity lay in finding a way to simplify this technology for the mass market, thereby making it more accessible.

Perhaps, in our previous letter, we failed to articulate what this potential solution might look like. Our thoughts were leaning towards an easy-to-use application, something along the lines of Microsoft Excel or Word.

At the end of that letter, we included a substantial excerpt from an interview with NVIDIA's CEO, Jensen Huang. Looking back, his words now seem eerily prophetic. He suggested, albeit vaguely, a solution to the complexity problem. He said:

*But finally, we have this piece of this new technology called artificial intelligence that can write that complex software so that we can automate it. The whole goal of writing software is to automate something. We're in this new world where, **over the next 10 years, we're going to see the automation of automation**".*

From the beginning, our core hypothesis has been that the 'intelligence' part of AI – the magic happening behind the scenes – would be incredibly beneficial to humanity. What we did not anticipate was how we would interact with this intelligence.

Human-like Manner to Use AI Technology

We must confess, for us, the ChatGPT product was also a revelation. In hindsight, it's quite an obvious solution. Whilst we were not entirely correct about the interactive form of AI technology, we were closer to the problem. We think a substantial bottleneck has been solved through large language models (LLMs) with a technique called Self-Attention. *Why?*

ChatGPT has quickly become one of the most successful consumer products in history. We attribute this, at least in part, to its low-friction method of human-to-machine interaction - or Human-Machine-Interface (HMI). The prompt-response style of interaction is nothing short of brilliant. Upon our first interaction with it, the value proposition was immediately clear.

It's often the case that breakthrough technologies gain traction due to advancements in HMIs, relieving users of the burden of complexity. Historically, humans have needed an interface to translate our thoughts into a language that machines can understand - binary code. This 'translation' problem between humans and computers has meant that communication is painfully slow, on the order of 400 characters a minute. By contrast, computers can communicate with each other up to a billion times faster². This discrepancy has traditionally meant that inputs and outputs have been rather linear.

Since the inception of the personal computer, technologists have been crafting small yet significant solutions to extract usefulness from these machines:

- *Graphical User Interface*: In the early 90s, the graphical user interface (GUI) was the primary tool enabling communication between humans and personal computers. It's noteworthy that Bill Gates recently penned an article entitled "*The Age of AI Has Begun*"³. In it, he paralleled a 2022 demonstration by OpenAI of their ChatGPT to the seminal moment in 1980 when he was first introduced to a Graphical User Interface (GUI). Gates posed a challenge to the team - for the LLM to pass a 60-question college-level biology exam (it correctly answered 59 questions).
- *Fingers and Touch*: When Steve Jobs unveiled the iPhone in 2007, he spent the initial nine minutes of his presentation discussing the limitations of human interaction with mobile devices. At that time every mainstream smartphone in the market had a physical keyboard (restricting screen space) or a stylus (easily misplaced) as an HMI. Apple's revolutionary idea was to introduce a full screen and eliminate the keyboard by utilising multi-touch.
- *Apps*: Nowadays, we use countless apps, essentially graphical interfaces, which translate our specific tasks into computer language. Given their specialised functions, we have millions of different apps tailored for diverse use cases.

The success of LLMs⁴ can be attributed to their significant shift away from the need for physical and software tools for eliciting intelligent responses from machines. These models allow us to converse in human terms, using the most fundamental tool of our

² at 3,200 bits/s and the latest networks communicate at 3.44 Pbps or 400 GB/s

³ Source: <https://www.gatesnotes.com/The-Age-of-AI-Has-Begun>

⁴ Whilst we think the other nodes of Generative AI (GANs, RNNs and diffusion models) are interesting, we think the core Large Language Transformer model that moves the technology closer to Artificial General Intelligence will be the most impactful.

communication - *language*. More importantly, these LLMs provide considerable value by generating in-depth, insightful outputs in response to a few simple prompts.

Information Leverage

The crux of AI's strength lies not in amplifying physical capabilities, as technologies like the automobile, aeroplane, or bicycle have done, but in its unique capacity to exponentially enhance our information leverage.

Here is what we mean: consider the sheer volume of accurate, pertinent, and relevant information one can extract from ChatGPT by simply asking a few questions.

The internet is teeming with instances of people creating games, coding, poems, and even philosophical thinking from a few crisp prompts. This mirrors the non-linear relationship automobiles brought us, linking limited input (fuel and an engine) to remarkable output (speed and distance).

LLMs represent a nascent version of general artificial intelligence, heralding a multi-decade installation and deployment period of 'information leverage'.

The crucial point for us is that today, LLMs have served a purpose beyond mere speculation, igniting the collective imagination, and shedding light on the myriad of possibilities.

Now, 'Everyone is a Programmer'

Our opinion is that this AI epoch will precipitate profound upheaval at the macro level in many societal systems (see later). As the diffusion of technology progresses, several industries are ripe for change. Consider how the Internet commoditised the cost of the printed word and altered the distribution model - disrupting the physical bookstore industry.

Controversially, we envision AI as the catalyst for a reconfiguration of software development.

At this year's GTC, Jensen Huang shared an insightful point:

*"Generative AI is a new kind of computer, one that we program in human language. This ability has profound implications. **Everyone can direct a computer to solve problems.** This was a domain only for computer programmers. **Now, everyone is a programmer.**"*

Our observation is that most early adopters of ChatGPT often view it more as a novelty rather than a useful tool. It can create novel outputs such as a poem on any subject or provide answers to general knowledge questions.

We believe that AI's initial impact will be to democratise coding, reduce barriers and broaden participation in software development. Despite an estimated 30 million developers worldwide, only a small percentage of the global population possesses the skills to create software for the billions of devices in existence.

Approaching the issue from a first-principles perspective, we question the necessity of GUI apps when we can simply pose a question (through a prompt or voice). Why should we navigate dropdown boxes, menu items, and buttons to find the information we need? Is there a requirement for separate task apps, payroll apps, and CRM apps?

To be clear, we are not calling for the immediate death of application software development. As we will discuss later, a technology diffusion period can span 2–3 decades.

Over this shifting landscape, we anticipate a transformation of the entire software stack, ranging from the operating system up to the application. This transformation will offer an incredible investable opportunity for us.

Beyond this obvious impact, technological innovations at the 'epoch' level extend their reach beyond niche sectors; they orchestrate profound societal changes.

We have endeavoured to articulate our thoughts on what this could potentially look like.

Technology Epochs = Seismic Shifts Economically, Politically, and Socially.

Reflecting upon the epochal moments like the Wright Brothers' triumph at Le Mans or Ford's revolutionary assembly line, we reckon there were a few nerds (like us), expressing optimism that perhaps bordered on hyperbole. However, for society at large, these breakthroughs were inconsequential to their immediate lives because the technology required a whole installation of underlying infrastructure financed with risk capital.

As investors, it's crucial for us to comprehend that technological epochs are slow-burning revolutions spanning four to six decades. These epochs are punctuated by shifts in technology, industry trends, social norms, legal frameworks, and economic paradigms.

Predicting with precision what the fifth decade could look like from our vantage point in the first is stepping into the realm of science fiction authors (to be fair there is an argument of thought that science fiction can be highly prophetic). But there appear to be some common 'rhymes' with each technology epoch. From an investor's perspective, we could at least recognise the 'seasons' to come, even though we could not assimilate the exact weather prediction. We could prepare to bring our umbrellas in summer and warm clothes in winter, so to speak.

In the face of these unknowns, one might well ask, "*How is an investor to find his or her way?*" To help answer that question, I find myself frequently reaching for a book that's been well-thumbed through over the years, Carlota Perez's "*Technological Revolutions and Techno-Economic Paradigms*"⁵.

Perez's book is quite the journey through time, taking us through a host of technological revolutions: the Industrial Revolution to the Age of Steam and Railways, the Age of Steel, to Oil and Mass Production, and finally landing us in the Age of Information Technology.

Each of these epochs, it's worth noting, has spanned about 30 to 50 years before yielding to the next. What's particularly fascinating is that Perez pegs the dawn of the last epoch, the Information Age, at around 1971. Now, if you do a bit of mental arithmetic, you might find yourself raising an eyebrow at the prospect of us now being on the cusp of a new "Age of AI" epoch in 2023.

⁵ Source for all Perez quotes across this letter: [link](#)

Is AI a New Technology Epoch?

Perez articulates technology revolutions (or epochs) as:

“A technological revolution can be defined as a powerful and highly visible cluster of new and dynamic technologies, products and inspires, capable about bringing about an upheaval in the whole fabric of the economy and of propelling a long-term upsurge of development”.

In navigating the economic implications of these paradigm shifts, she notes that *“technological transformation brings with it a major shift in the relative price structure that guides economic agents towards the intensive use of the more powerful new inputs and technologies”.*

Two common threads run through all these epochs:

- the emergence of a technical breakthrough, and
- what Perez terms an 'attractor,' or a force that renders the innovation economical or introduces business innovations that prove cost-effective.

In short, an attractor provides a solid business case for society to embrace technology.

We've put forth the idea that 'information leverage' might fit the bill as an attractor. To underline its significance, consider it against a recent technological development like cryptocurrency, which currently lacks a significant business case.

Major technological epochs spur widespread change. First, they generate new products and give birth to adjacent technologies and industries. Second, they bring about technological advancements within existing mature industries, leading to a reshuffling of winners and losers⁶.

A Mass Reconfiguration

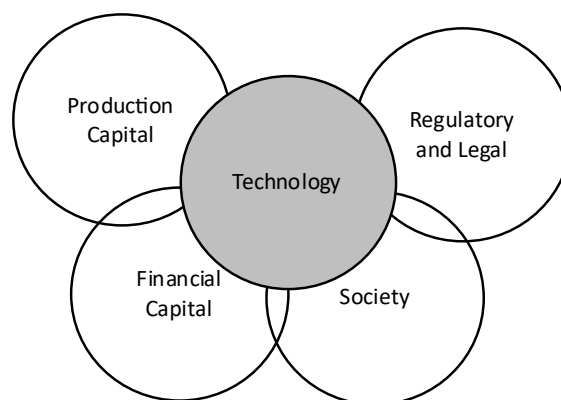


Figure 1- Systems Impacted by technological revolutions (SaltLight Research).

⁶ Clayton Christensen, a leading thinker on this subject, has written extensively on how supply chains can be disrupted by new technologies. He observed how attractive profits in one part of the supply chain can evaporate and mushroom elsewhere due to a process he called modularity and commoditisation. By way of example: the newspaper business was great (a significant investment area for Warren Buffett) until the Internet came along.

Perez's book lays out the 'growing stages' that these technological epochs tend to follow. It's critical to understand that these technologies don't exist in a vacuum. Their diffusion is an intricate dance that takes place within a web of social, legal, cultural, and capital paradigms—all of which must be rearranged to accommodate the newcomer.

Think about how the advent of the automobile transformed our cities, giving rise to the suburban lifestyle. Consider how the industrial revolution created employment for large corporate organisations. Reflect on how the railway system revolutionised regional communications, shrinking distances and accelerating the pace of life.

The Road to Revolution

What we find most intriguing in Perez's work are her insights into the potential pathways for the diffusion of technology. Interwoven with these possibilities are the lessons we've gleaned from history regarding the role of financial capital. The past can be a powerful teacher if we're willing to listen⁷

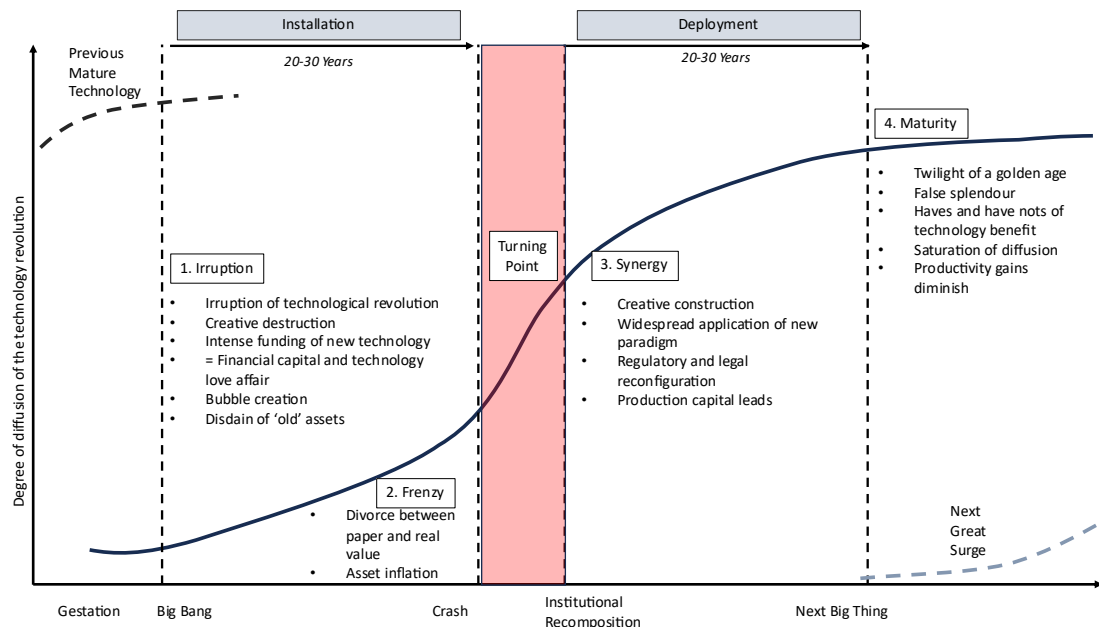


Figure 2 - Degree of diffusion of a technology revolution over time (Source: Perez)

Perez neatly dissects the technology adoption curve into an "installation period" that she says: "ploughs ahead like a bulldozer, disrupting the established fabric and constructing the new industrial networks," and a subsequent "deployment period".

- The installation period itself is broken down into "irruption" and "frenzy" phases.
- Irruption Phase:* Crucially, during the installation period, while the technology and financial capital are expanding at breakneck speed, there's a discord between the existing societal and regulatory systems. This friction slows

⁷ Perez makes the distinction between financial capital and production capital. Financial capital is considered "making money from money" whereas production capital involves agents who generate wealth from providing goods and services (retained earnings capital).

down the pace of adoption. The necessary externalities such as infrastructure and distribution mechanisms still need to be developed.

- *Frenzy phase*: The impact of these technological revolutions on capitalism is readily apparent. Industries crumble under the weight of Schumpeter's creative destruction. During the frenzy phase, financial capital often gets carried away, growing overly optimistic about prospects before the system has fully adapted to the new technology. This over-optimism leads to boom-and-bust cycles, exemplified by the dot.com crash, where many promising ideas were simply too premature.
- *Turning point*: Eventually, the installation period crescendos with a '**turning point**', a pivotal juncture often marked by a severe recession, which necessitates a reconfiguration of the socio-economic and regulatory systems.
- Once these elements are realigned, growth resumes during the "**deployment period**".
- *Synergy phase*: This period is divided into a '**synergy**' phase when the entire system has been recalibrated to accommodate the new technology.
 - Society has adapted, and the early casualties of disruption have been laid to rest. The early externalities have now been deployed and installed across society (think of the construction of highways nationwide to facilitate the full utilisation of automobiles).
 - This leads to a golden era of production capital generation. Businesses that have adapted, both old and new, are turning profits and accumulating retained earnings. Perez describes it as "*a time of promise, work, and hope. For many, the future looks bright*".
 - A prime example is Amazon, born two decades into the information revolution. Despite nearly being wiped out by the dot.com crash, Amazon survived thanks to an exceptional management team. Nevertheless, the company needed a robust delivery infrastructure and the widespread adoption of e-commerce to truly realise its economic value.
- *Maturity*: Finally, there is a '**maturity**' phase that heralds the end of the golden age, although it shines with deceptive brilliance. Markets gradually reach saturation.
 - Product life cycles shorten due to the widespread accumulation of knowledge on how to create the technology. Perez notes, "*Gradually the paradigm is taken to its ultimate consequences until it shows up its limitations*".
 - A case in point is the mobile handset market, which is saturated today. Companies are resorting to monopolistic actions to sustain growth. Social media opportunities are about gaining market share and captivating the lowest common denominator of attention. Perez observes, "*But the unfulfilled promises had been piling up, while most people nurtured the expectation of personal and social advance. The result is an increasing socio-political split*".

The Artificial Intelligence Epoch

Forecasting the course of AI's evolution feels a bit like juggling flaming torches while blindfolded. If you're lucky and skilled, you might put on quite the spectacle; but there's always the risk you end up with scorched trousers. Nevertheless, we're going to dare to guess, with the hope that our intellectual trousers remain intact.

We're likely in the early 'irruption' phase of AI, which is gradually taking over from the maturity phase of the information technology era (which, according to Perez, began in 1971 and, we speculate, culminated with the rise of the FAANG businesses). AI technology holds immense potential, but it's also grappling with typical early-stage limitations: speed, data age, and crucially, cost.

It'll be for historians to pinpoint when the AI epoch began, but we can confidently say that the infatuation with AI is just getting started. We predict that LLMs will quickly become a generalised base layer. The more captivating applications, and by extension, a broader range of investable opportunities, are likely to arise from domain-specific models.

Models like OpenAI's ChatGPT, Google's PaLM, Meta's LLaMA, and others on the horizon, will likely serve as foundational models. These generalised language AI models, trained on public Internet data, will facilitate basic language interactions. The nuances between them will depend largely on proprietary datasets. They'll become a layer of AI infrastructure, akin to cloud IaaS services, on which AI applications will be built. However, a significant challenge lies in the fact that these models lack access to your domain-specific data. Ingesting and learning that data currently presents a classic data size problem.

This harks back to the early days of the personal computer when storage and memory were significant bottlenecks. LLMs are currently facing a similar bottleneck, measured in tokens⁸, which limits the amount of information they can ingest (approximately 819 words, given current computational and memory limitations). The workaround—breaking a document into multiple queries—is both inefficient and costly in terms of API calls. However, we can confidently anticipate that these issues will be resolved as computational and memory technologies continue to evolve.

Once these obstacles are overcome, we can envisage an explosion in enterprise applications of AI.

What Does This Mean for Investors?

Assessing the investment landscape at this juncture, the publicly traded AI opportunities appear to be somewhat scant. Our portfolio has held NVIDIA for a considerable time. While there are a select few AI-specific public beneficiaries, their scarcity tends to suggest that we are still in the 'irruption' phase. Underneath the surface, there's a veritable seismic activity of venture capital-backed AI start-ups, each backed by a noteworthy (albeit, in comparison to consumer tech funding, modest) influx of financial capital.

Some market participants hold the belief that the incumbent tech giants such as Google, Microsoft, and Facebook, are poised to be the major beneficiaries of the AI

⁸ A Token is a discrete unit of text (not necessarily a fully word, but a word or character) that is encoded to form the space of content for the model to work out context and answer questions.

epoch. These companies, with their extensive teams of data scientists, expansive distribution networks, and hefty capital resources, seem to be in a prime position to seize a significant slice of the value to be derived from AI.

But if we take a gander at historical precedents, it's far from a foregone conclusion that the most resource-rich players will emerge victorious. In past technological revolutions, it wasn't always the giants who leveraged the emerging technology to create substantial value.

As the Wright Brothers were working in their garage, Samuel Pierpont Langley, a prominent scientist, and aviation enthusiast was being supported by the US government and the Smithsonian Institution. Despite Langley's advantages, it was the Wright Brothers' unwavering determination and resourcefulness that propelled them to achieve the impossible.

Much of the automobile industry was focused on electric power drivetrains. Indeed, Henry Ford, while employed at an electric vehicle company, spent his nights working on the internal combustion engine (ICE), ultimately changing the course of the industry.

We leave open a significant window that any early success of incumbents could be disrupted by a nimbler misfit. We're also open to the possibility of non-tech incumbents who can capitalise on AI technology in a 'sustaining' innovation that augments their existing business.

Old Investment Principles Apply

It is essential to remember that, amidst the hype surrounding new technologies, investors should not disregard the time-tested principles of rational investing. Even in the realm of AI, companies must possess a competitive advantage, the potential for high returns on capital, and a unique product offering. However, it's worth noting that profits are more likely to emerge during the 'deployment' stage rather than the speculative 'installation' stage.

Consequently, at this early juncture, we find meagre enticement in wagering bets solely on AI models. Our conviction remains that such models can be somewhat effortlessly commoditised. As we've underlined previously, the genuine potential resides in proprietary data, as it may provide a possible competitive advantage for corporations navigating the AI territory.

Reflecting on the Internet Age, we witnessed the aggregation of distribution coupled with a pivotal 'connector' role that generated substantial value for shareholders. Concepts such as network effects, economies of scale, and control over a scarce resource will serve as the differentiating factors separating the 'good' AI investments from the 'bad'.

Selling Shovels in the Irruption Phase

Given the wide range of outcomes at the irruption phase, we have adopted a prudent approach reminiscent of the savvy individuals who profited during the gold rush not by mining for gold themselves, but by selling the essential shovels. Picking winners and losers in a state of flux is challenging.

We have directed our focus towards companies operating within the semiconductor supply chain and those companies that prepare the data for model training. In

semiconductors, we recognise the scarcity of entities capable of manufacturing the requisite computational capacity. Moreover, we observe a formidable barrier preventing AI companies from vertically integrating into the more intricate aspects of the manufacturing process.

The Expense of AI Compute

Currently, one of the primary challenges facing the field of AI, when juxtaposed with services such as search, lies in the remarkable resource intensiveness of language model (LLM) computation. These LLMs comprise billions of parameters and necessitate copious amounts of data, computing power, and energy for training. Numerous studies have substantiated the significance of size in this context.

The considerable capital outlay required for training a model is truly staggering; as an illustration, Dylan Patel of SemiAnalysis⁹ estimates that the cost of training a model with 1 trillion parameters would amount to approximately \$300 million. This initial investment of substantial magnitude poses a hindrance to the regular updating of the model with new information, thereby elucidating why the knowledge of ChatGPT 3.5 is confined to developments up until September 2021.

Furthermore, the inference process, wherein end-users derive value from the trained model, is also fraught with expense. Analysts suggest that an LLM query costs 2-3 times more than a general search query. Google's initial advantage, which propelled its ascendance, was rooted in its astute utilisation of commoditised PCs, as opposed to its competitors' reliance on high-end server capacity. This cost advantage afforded Google the ability to reinvest profits into enhancing results, capturing a larger user base, and ultimately surpassing the competition.

It is pertinent to note that the industry tends to respond to cost challenges of this nature by developing new technologies over time. We do not harbour any doubts regarding the eventual substantial reduction of these costs. However, we struggle to see the economics compare favourably to a typical CPU-based search query.

AI and the Innovator's Dilemma

When it comes to scaling up to address the vast search market, AI will not offer a uniform solution for every business. Companies like Google find themselves facing a classic Innovator's Dilemma. Search, traditionally monetised through advertising, encounters a challenge with the introduction of LLM queries, as the revenue generated from ads is unlikely to yield the same margins as before. This margin predicament raises a crucial question: should one restrict the results of LLMs or accept lower margins?

OpenAI has pursued an alternative solution by implementing a subscription model. However, for Google, this poses the risk of cannibalising their existing search product and potentially excluding lower-income users from accessing the service.

The Role of Hardware in AI

Presently, the true bottleneck in AI lies in compute capacity. NVIDIA, a company that has been diligently preparing for this moment for the past five years, is currently the dominant player in the field. Through organic development and strategic mergers and

⁹ Source: <https://www.semianalysis.com/p/the-inference-cost-of-search-disruption>

acquisitions, NVIDIA has been laying the foundation for addressing this challenge. Recognising that efficiently moving bits within these large models would be a critical obstacle, NVIDIA acquired Mellanox in 2020.

In our investment approach, we have analysed the supply chain to identify opportunities that contribute to solving the AI compute bottleneck. By working our way up the supply chain, we have arrived at ASML, a company that manufactures the tools necessary for producing the semiconductors that will alleviate this compute bottleneck. By investing in ASML, we position ourselves to benefit from their crucial role in enabling not only NVIDIA but also other potential entrants in the AI space.

Inevitability: The Frenzy Stage and Subsequent Boom and Bust

We conclude with a long-term caution to ourselves. As we observe the progression of AI technology, we hypothesise that once it reaches a level of proficiency to disrupt knowledge-based human resources, the economic metrics used to gauge its impact will revolve around cost savings on employment expenses. Just as the Internet was once evaluated based on the number of eyeballs, AI will be measured by the number of jobs it saves.

As the “frenzy” stage progresses, the inherent potential baked into valuations and the staggering volume of financial capital being thrown about could start to overshadow the hard realities of AI. It strikes us as rather like the enthusiastic dog chasing after a car – the beast doesn't pause to consider what it might do if it ever caught up with the vehicle. A course correction will, in due course, become necessary if we're to truly squeeze the juice out of this technology.

When AI technology begins to take a meaningful bite out of the labour market, it's going to become painfully apparent that our societal, legal, regulatory, and economic scaffolding isn't quite cut out for an AI-dominated world. It's like trying to retrofit a square peg into a round hole – something's bound to give eventually.

In light of these patterns, it wouldn't surprise us to see a so-called “turning point” recession where the mechanisms of our economy and society are forced to realign in response to the pervasive reach of AI technology. It's an uncomfortable truth, but these reckonings tend to act as a necessary pruning process.

The shape of this looming transformation, we confess, is far from clear. Economic output, after all, stems from the twin fountains of productivity and population growth. This leaves us to ponder the puzzle of how GDP might be gauged in a society capable of cranking out AI entities that not only match but could potentially leave human productivity in the dust.

It's a thorny issue, to say the least, akin to trying to herd cats or nail jelly to the wall. Difficult questions will need to be answered around the distribution of capital and labour. These very conundrums echoed throughout the dying days of the industrial and mass production eras and they're far from settled even now.

Most of our rambling discourse here has dwelled on the broader implications of artificial general intelligence (AGI), but we haven't even begun to scratch the surface of other AI applications like autonomous vehicles, factory automation, and robotics. That, however, is a rabbit hole for another day. There's only so much intellectual hay we can make in a single sitting, after all. The remaining fodder must be left for a future conversation.

Feel entirely welcome to challenge our perspectives and append anything you feel we may have overlooked. We've taken the trouble to pen these thoughts in part because it helps shape our cognitive map as we peer into the hazy distance of our 2028 time horizons. We're under no illusions that some of our musings may end up dissipating in the wind, much like the contents of previous letters. Above all, we desire to correct our course when we've erred - we have no room for pride in our pursuit of truth.

As always, we take this moment to remind our co-investors that the lion's share of our liquid wealth is invested right alongside yours in the same fund. We are on this ship together, riding the same crests and troughs of the market. We are with you in the profits and, just as inevitably, the losses. And that's how it ought to be, we would not have it any other way.

Warmly,

David Eborall

Portfolio Manager

Disclaimer

Collective investment schemes are generally medium to long-term investments. The value of participatory interest (units) or the investment may go down as well as up. Past performance is not necessarily a guide to future performance. Collective investment schemes are traded at ruling prices and can engage in borrowing and scrip lending. A Schedule of fees and charges and maximum commissions, as well as a detailed description of how performance fees are calculated and applied, is available on request from Sanne Management Company (RF) (Pty) Ltd ("Manager"). The Manager does not provide any guarantee in respect to the capital or the return of the portfolio. The Manager may close the portfolio to new investors to manage it efficiently according to its mandate. The Manager ensures fair treatment of investors by not offering preferential fees or liquidity terms to any investor within the same strategy. The Manager is registered and approved by the Financial Sector Conduct Authority under CISCA. The Manager retains full legal responsibility for the portfolio. FirstRand Bank Limited is the appointed trustee. SaltLight Capital Management (Pty) Ltd, FSP No. 48286, is authorised under the Financial Advisory and Intermediary Services Act 37 of 2002 to render investment management services.