Chips, Clouds, and Chimeras: Decoding the China Tech Clash

Tech Revolution Summit 2023
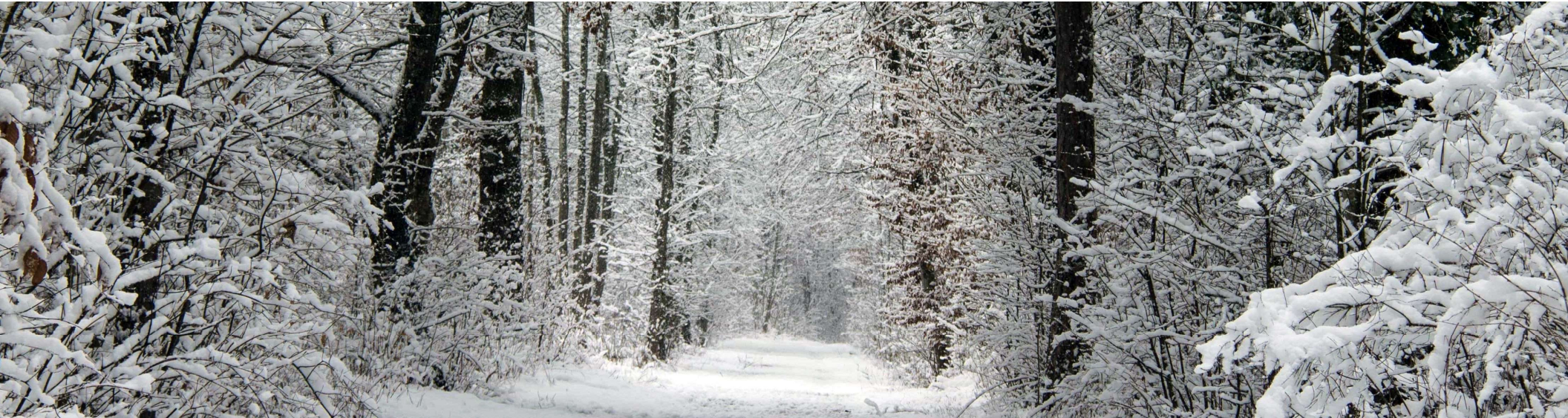
Dr. Richard Windsor PhD CFA
Radio Free Mobile

Fixed: +1 415 830 6032
Fixed: +44 203 239 9904
Mobile: +971 56 1142 482
Skype: rhswindsor
Twitter: @RadioFreeMobile

richard@radiofreemobile.com
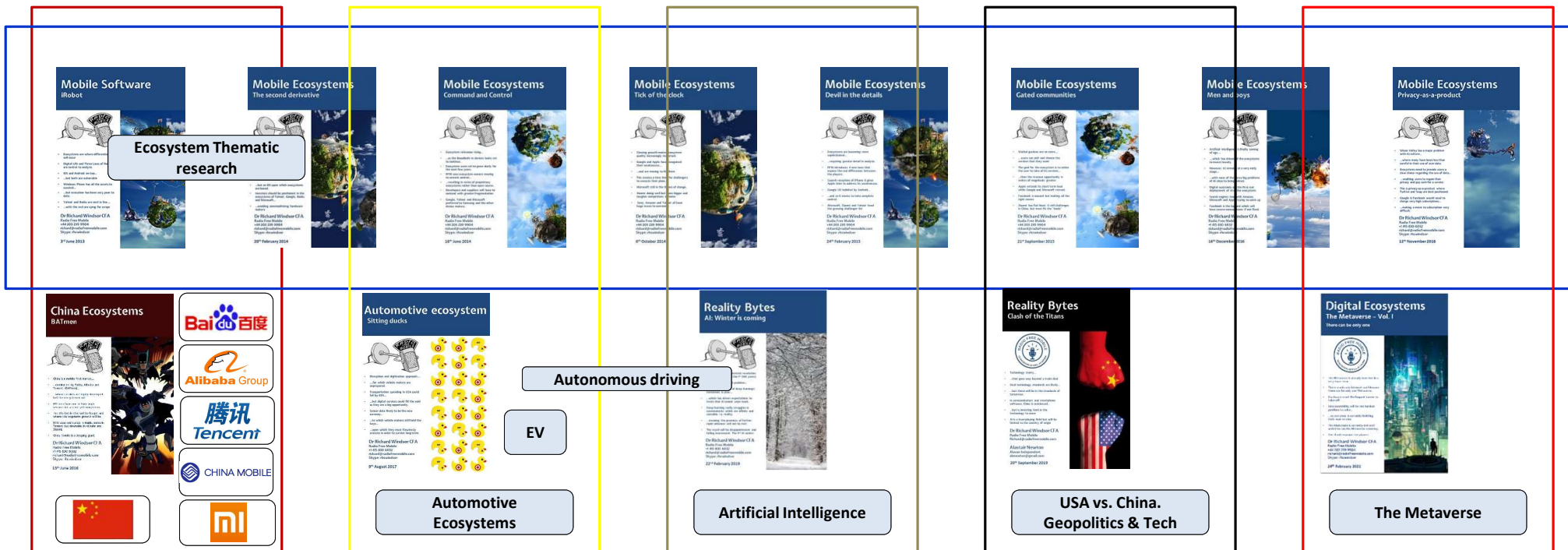www.radiofreemobile.com

November 2023
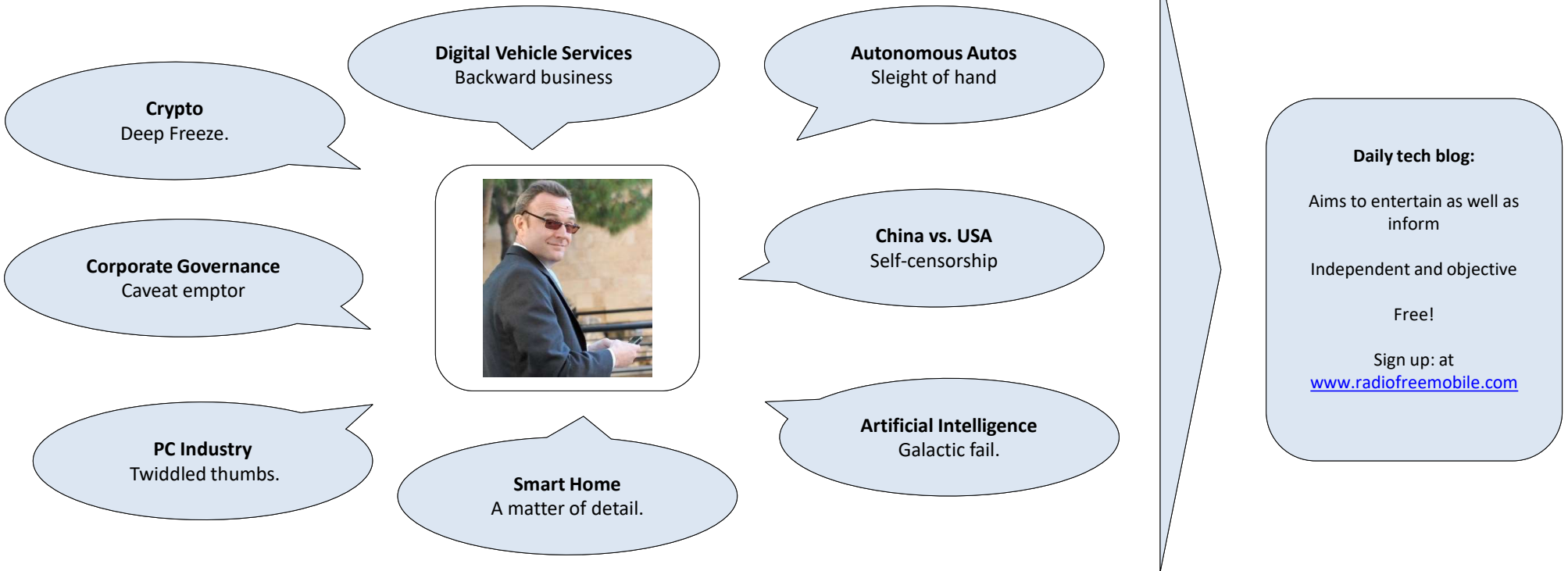
Radio Free Mobile

Independence and objectivity

The thematic research is complimented by in depth research on each individual ecosystems and themes.

Radio Free Mobile Research Coverage areas

Ecosystem Thematic research

Autonomous driving

EV

Automotive Ecosystems

Artificial Intelligence

USA vs. China. Geopolitics & Tech

The Metaverse

The daily blog is available at www.radiofreemobile.com

Radio Free Mobile Technology Blog

Crypto
Deep Freeze.

Digital Vehicle Services
Backward business

Autonomous Autos
Sleight of hand

Corporate Governance
Caveat emptor

China vs. USA
Self-censorship

PC Industry
Twiddled thumbs.

Smart Home
A matter of detail.

Artificial Intelligence
Galactic fail.

Daily tech blog:

Aims to entertain as well as inform

Independent and objective

Free!

Sign up: at
www.radiofreemobile.com

China vs. USA
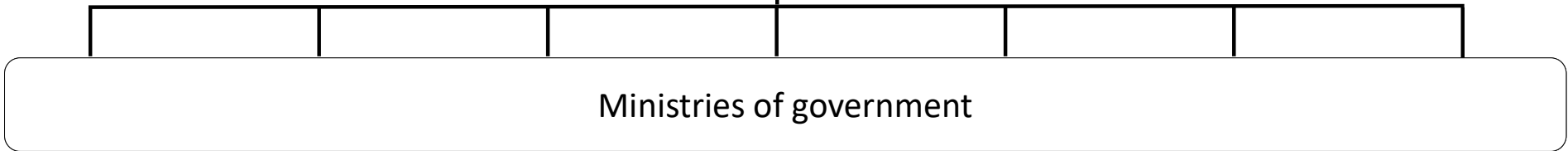
Clash of the Titans

RADIO FREE MOBILE
TO ENTERTAIN AS WELL AS INFORM
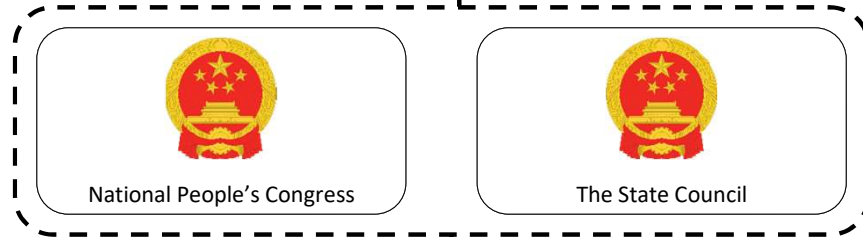
Politburo Standing Committee line-up makes it very unlikely that any of the current policies will change. This could mean doubling down on big policy blunders

The sheriffs of China
2023

The Politburo Standing Committee

Li Qiang

Promoted to No.2 and likely to be Premier.

This is despite the economic disaster of the 2022 Shanghai lockdown which Qiang oversaw.

Widely seen as a sign of the pre-eminence of loyalty over competence.

Little likelihood of easing of current policy

National People's Congress

The State Council

Ministries of government

Semiconductors

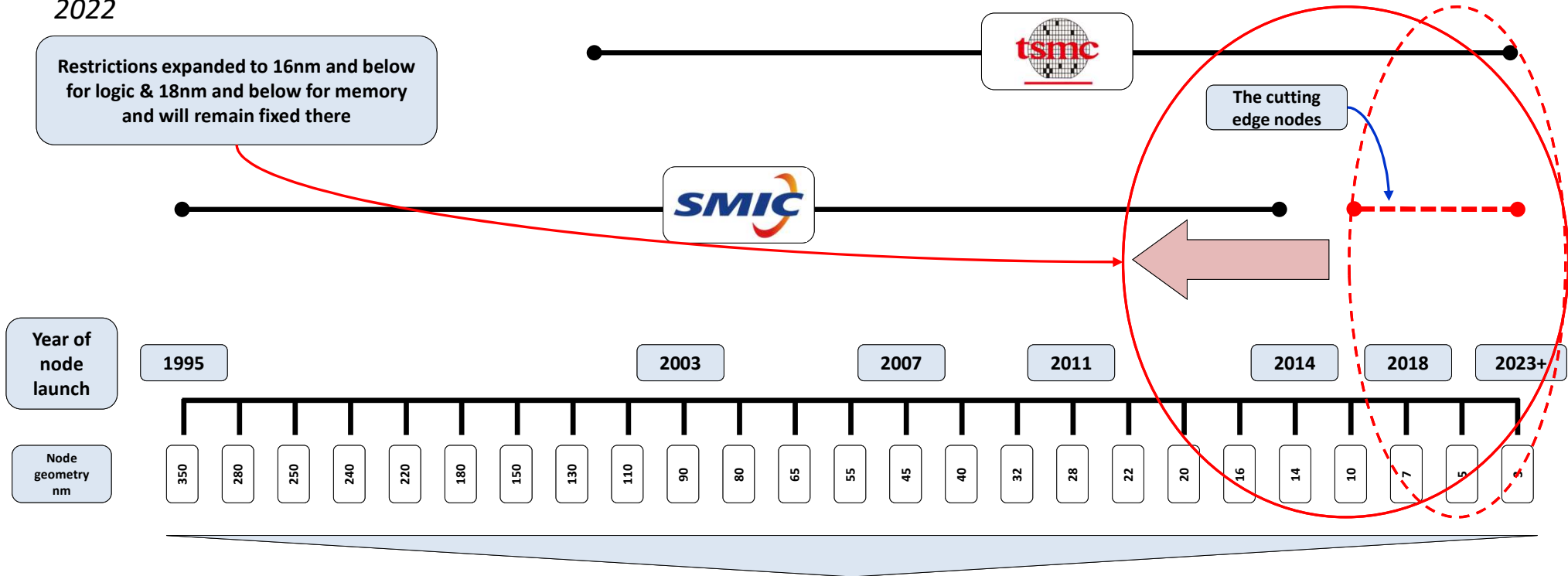The Netherlands and Japan join the party. January 23rd 2023

Ensures that China remains at least 10 years behind the USA as it has been for many years. The longer this persists, the further behind China will fall

Additional Export Controls

2022

Restrictions expanded to 16nm and below for logic & 18nm and below for memory and will remain fixed there

The cutting edge nodes

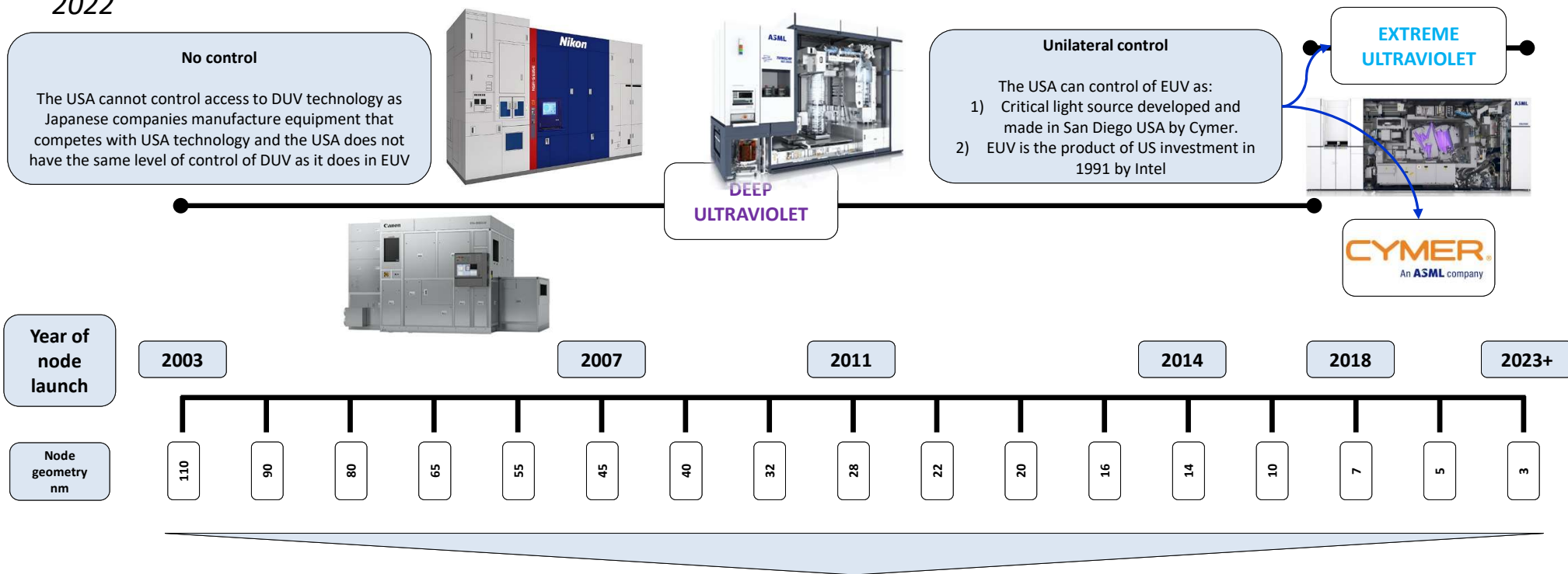	Year of node launch	1995								2003		2007		2011						2014			2018	2023+
Node geometry nm	350	280	250	240	220	180	150	130	110	90	80	65	55	45	40	32	28	22	20	16	14	10	7	5	3

A sliding restriction has become fixed. As TSMC et al move from 5nm to 3nm and beyond, the Chinese companies will further and further behind. This is a deliberate policy change to ensure that the USA remains as far as ahead of China as possible when it comes to semiconductor technology.

The USA need The Netherlands and Japan to impose restrictions in order to ensure the October 7th 2022 restrictions are not circumvented.

USA control of semiconductor technology by node
2022

No control

The USA cannot control access to DUV technology as Japanese companies manufacture equipment that competes with USA technology and the USA does not have the same level of control of DUV as it does in EUV

DEEP ULTRAVIOLET

Unilateral control

The USA can control of EUV as:
1) Critical light source developed and made in San Diego USA by Cymer.
2) EUV is the product of US investment in 1991 by Intel

EXTREME ULTRAVIOLET

Year of node launch	2003			2007							2014		2018		2023+

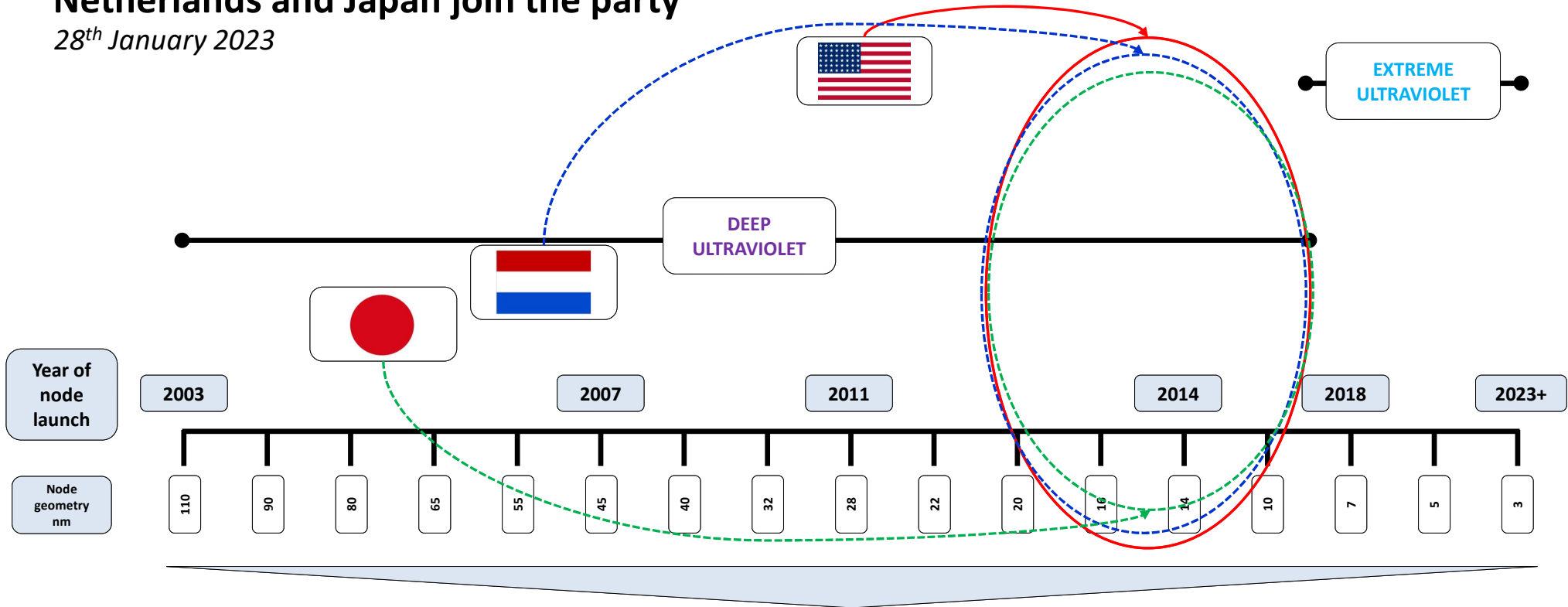| Node geometry nm	110	90	80	65	55	45	40	32	28	22	20	16	14	10	7	5	3

It is possible to build lagging edge fabs at 10nm - 18nm using non-USA technology. The US has been able to control leading edge through EUV, but needs the co-operation of Japan and The Netherlands in order to be sure that China will remain blocked at 16nm for logic and 18nm for memory.

Japan and The Netherlands have been less willing to impose restrictions as broad as the USA. Hence their restrictions are likely to be less severe.

Netherlands and Japan join the party

28th January 2023

EXTREME
ULTRAVIOLET

DEEP
ULTRAVIOLET

Year of node launch	2003				2007				2011			2014		2018	2023+

Node geometry nm	110	90	80	65	55	45	40	32	28	22	20	16	14	10	7	5	3

The USA appears to achieved most of its desires. RFM concludes: 1) The new restrictions will affect the non-leading edge nodes (18nm – 10nm), 2) China already has a lot of the equipment but will be unable to maintain or upgrade it. 3) Economic nodes will be mostly found at 28nm.

Huawei Mate 60 Pro+

2023, Propaganda in a phone

RFM is virtually certain that there is no 5G in this product.

Evidence is strong that SMIC have manufactured a 7nm SoC for Huawei.

However, RFM doubts that SMIC can manufacture millions of working chips at economic yields.

BUT: Semianalysis concludes that yields are high. Most evidence depends on data coming from China which could be propaganda. Pictures of gates are cleaner than would expect for low yield.

SMIC emulates TSMC and TSMC could not get this to work economically and shifted to 2 layers with EUV. RFM doubts that SMIC has succeeded where TSMC failed.

Conclusion

The lagging reality & Build the wall 2.0

China may now focus on the lagging nodes but whether these will be advanced enough to support the PLA is uncertain.
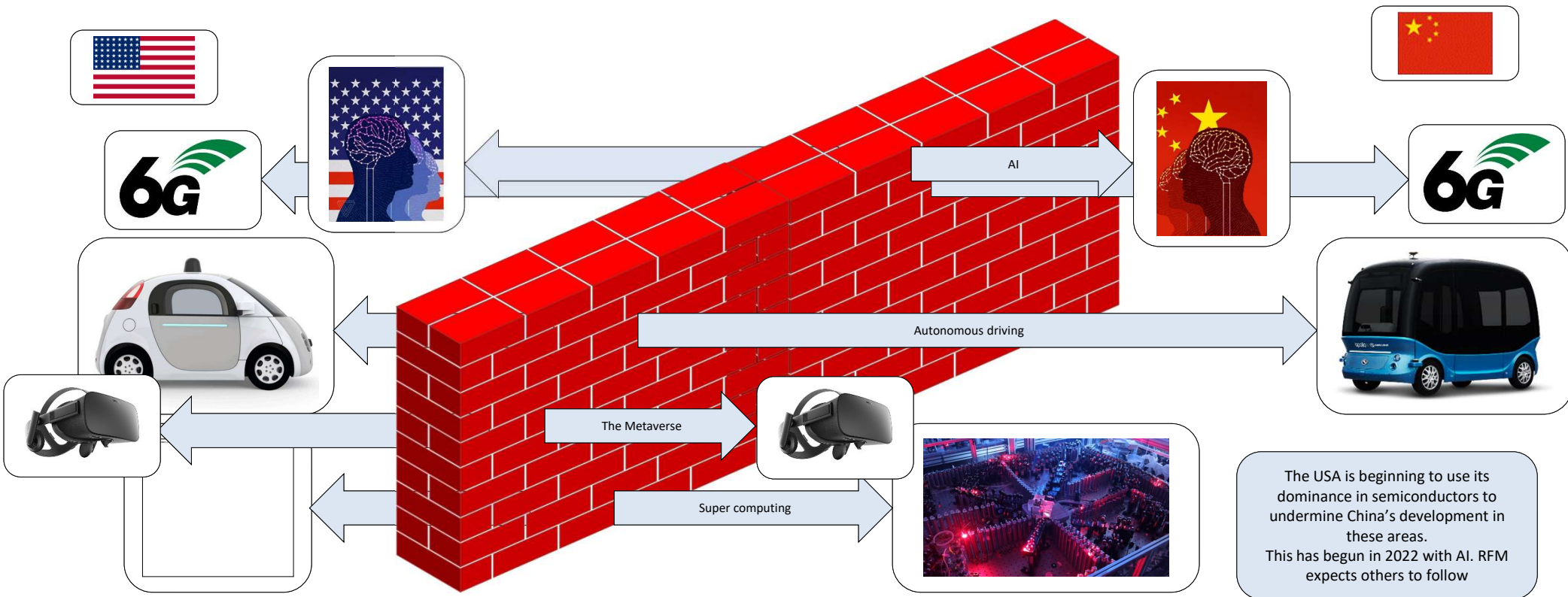
The lagging reality
2022

Where China is now expected to focus its efforts

The cutting edge nodes

Efforts in sub 20nm may now cease

Year of node launch	1995							2003			2007		2011				2014	2018	2023+
Node geometry nm	350	280	250	240	220	180	150	130	110	90	80	65	55	45	40	32	28	22	20	16	14	10	7	5	3
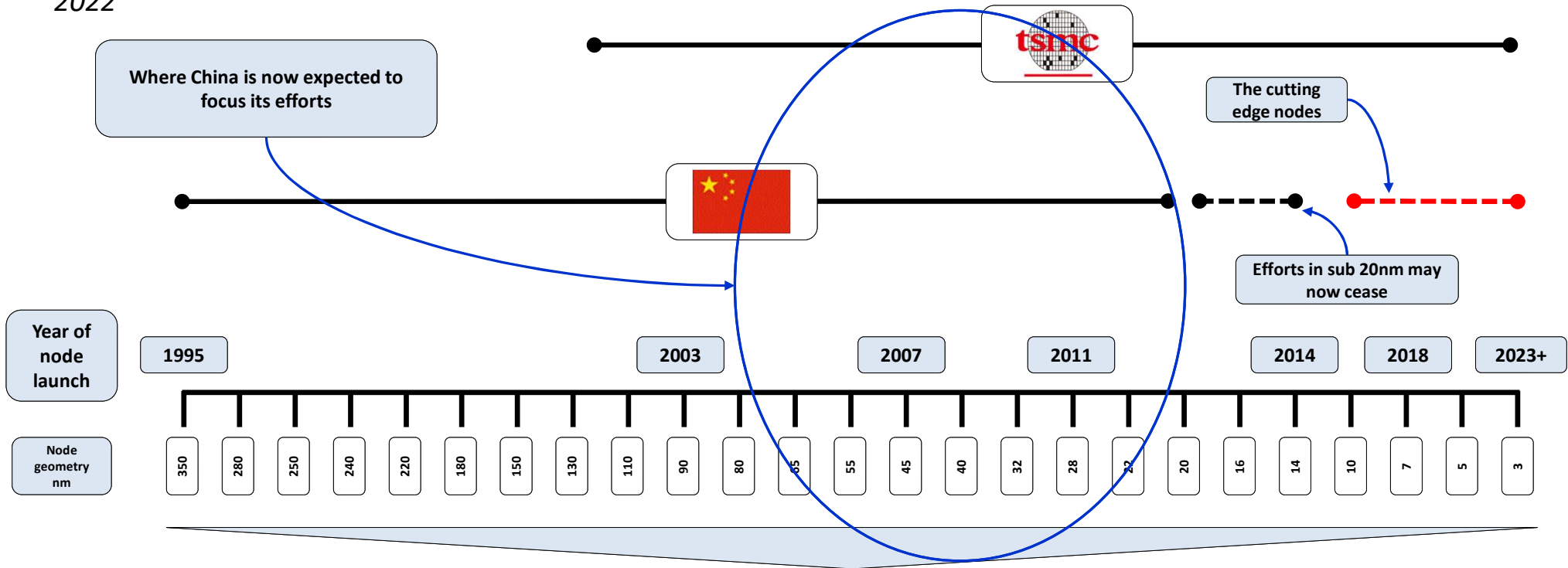
The 18nm nodes and anything below that are likely to prove so difficult to develop that the Chinese fabs may end up focusing on lagging edge nodes particularly 45nm to 28nm for which there is still plenty of demand. However as time passes, these nodes will fall further and further behind the leading edge which will continue to advance. Whether these will be enough for military and defence applications is very uncertain

www.radiofreemobile.com

It is the future standards where we see the wall being built. Established standards (like semis) are too hard / expensive to change
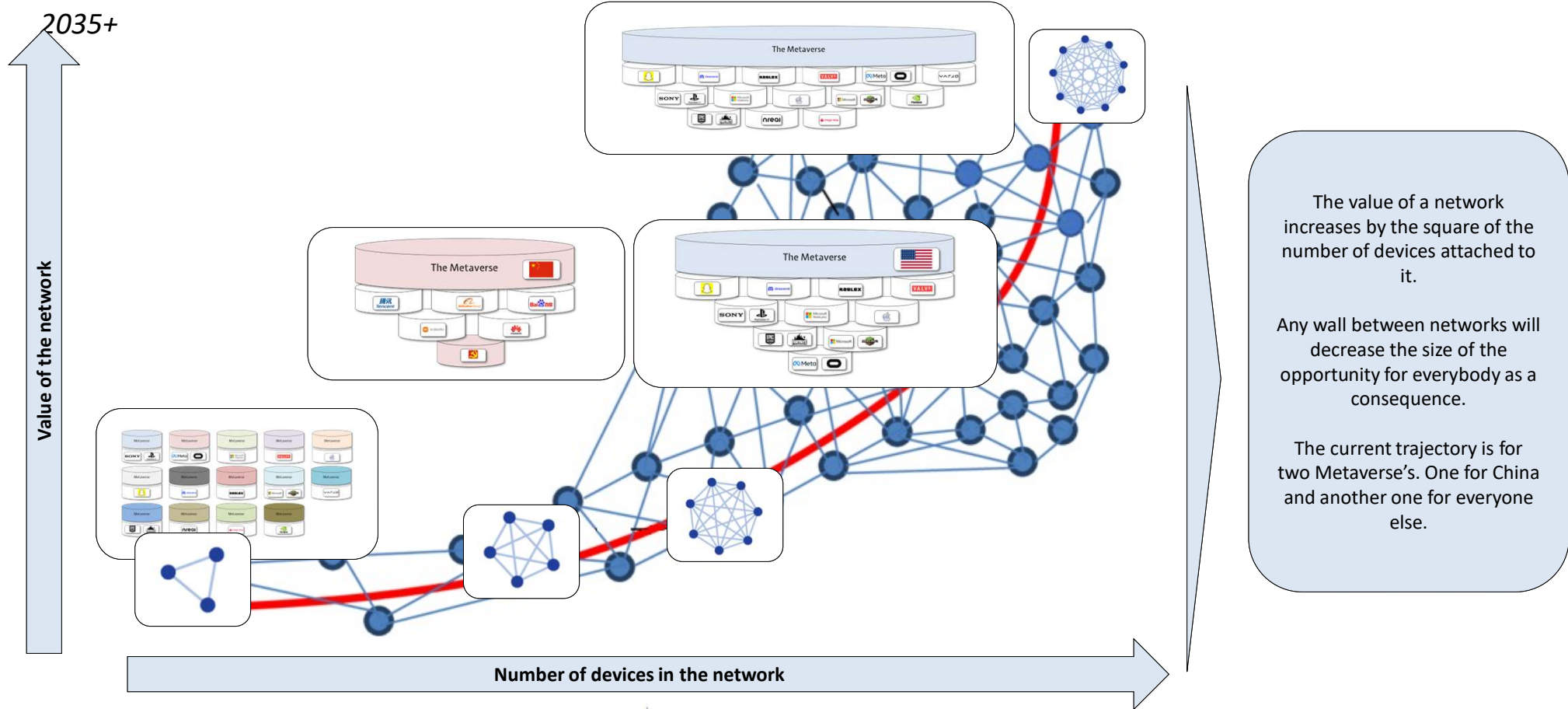
Technology war: future – Build The Wall!

AI

Autonomous driving

The Metaverse

Super computing

The USA is beginning to use its dominance in semiconductors to undermine China's development in these areas.
This has begun in 2022 with AI. RFM expects others to follow

The geopolitical trajectory remains towards a balkanisation of the Internet which will reduce the opportunity for all participants

The balkanisation of the Metaverse

2035+

Value of the network

Number of devices in the network

The value of a network increases by the square of the number of devices attached to it.

Any wall between networks will decrease the size of the opportunity for everybody as a consequence.

The current trajectory is for two Metaverse's. One for China and another one for everyone else.
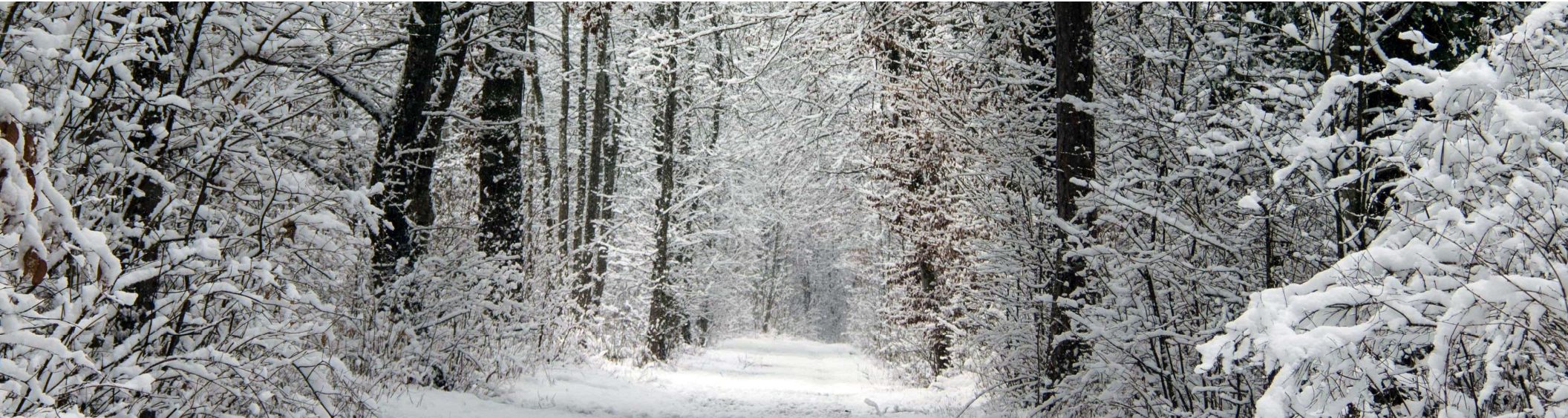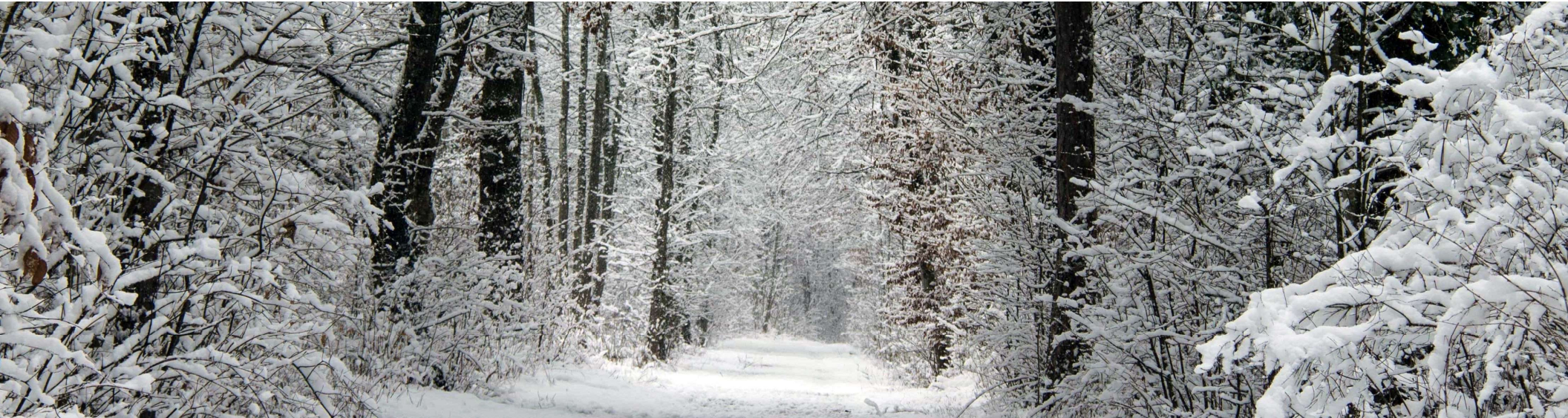
Artificial Intelligence

Bubble economics

The AI life cycle

Step forward not breakthrough

The current craze is a craze rather an a turning point because no real breakthrough or new technology has been developed.

AI Winter: past, present and future
2023

The signals Q2/Q3 2023?
(page 12)

Generative AI
rekindles hopes of
generalised AI

Machine
translation
turns out to be
slower, more
expensive and
inferior to
humans

Cold war
inspires
machine
translation
boom

Desktop
computers
become
cheaper and
better than Lisp
machines

Autonomous vehicles begin
to have difficulty.
Deep learning advocates
becoming quieter.
Visionaries return to
academia.

Alan Turing
publishes The
Imitation
Game

Digital assistants,
autonomous vehicles
bring AI into the general
consciousness. Hype
goes exponential

Lisp
Computing
launched

Backpropagation
technique invented

As hardware becomes more
powerful and data is
available backpropagation
begins to work in deep
learning

Reality re-asserted

AI sentiment, hype and investment

3

1

2

4

| 1950s | 1960s | 197 | 1980s | 1990s | 2000s | 2010s | 2020s | 2030s |

Fundamental characteristics

Unchanged for 40 years

Generative is not a breakthrough because it is using all of the old technologies. It is a big step forward in using language with machines.

# The ingredients of ImageNet and GPT-3
2012, 2020

2012

2020

2017

Backpropagation (1986)

Backpropagation (1986)

Vs.

GPUs

Google **TPUs**

Microsoft **$1bn**

Novel

Digital data

Digital data

Novel

IMAGENET

OpenAI GPT-3

There is no real novel change or proprietary technology that has been developed.

As such, all of the weaknesses that predate GPTs are still present in both GPT-3 and GPT-4
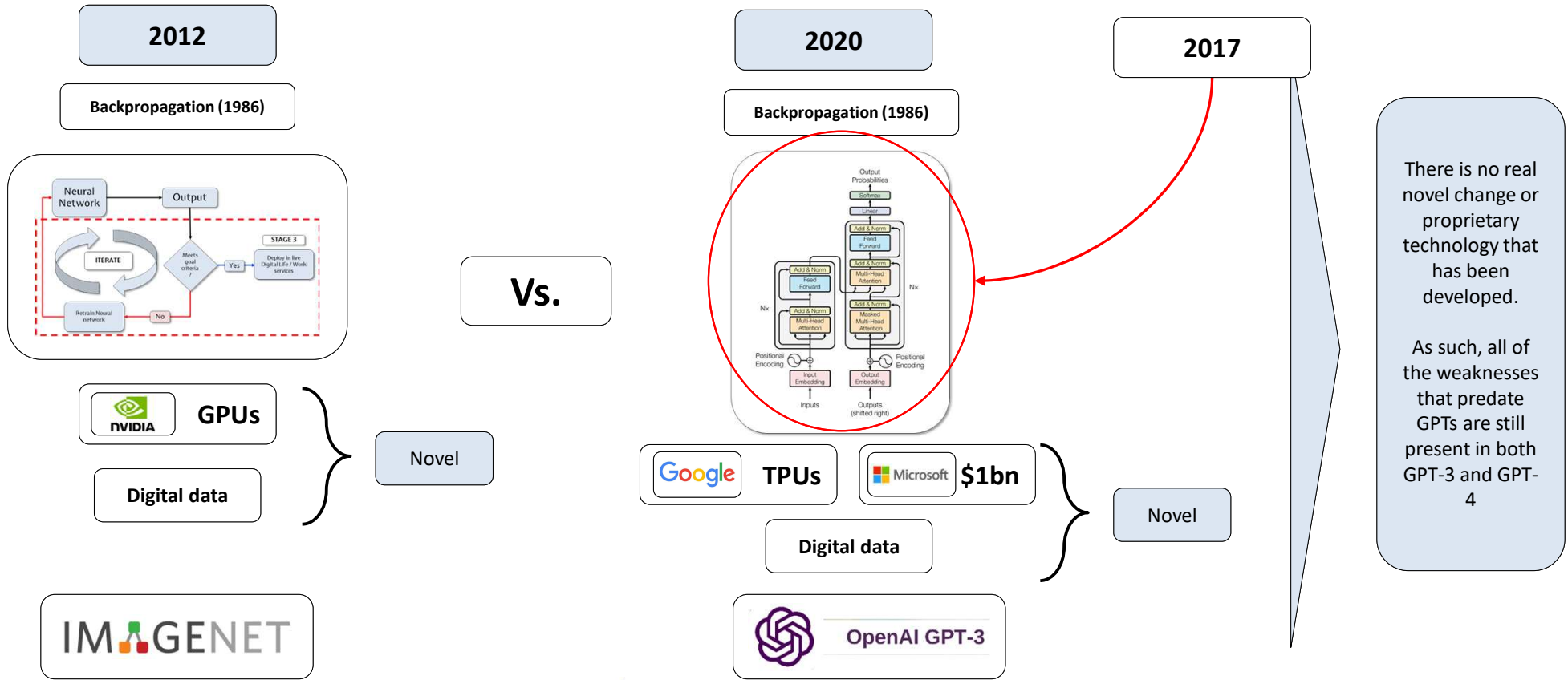
Deep learning only excels in situations and tasks that have both finite and stable data sets because it has no grasp of causality.

Deep learning's weaknesses: 1) causality
2015, 2016

Deep Mind's 2015 Atari Breakout demonstration
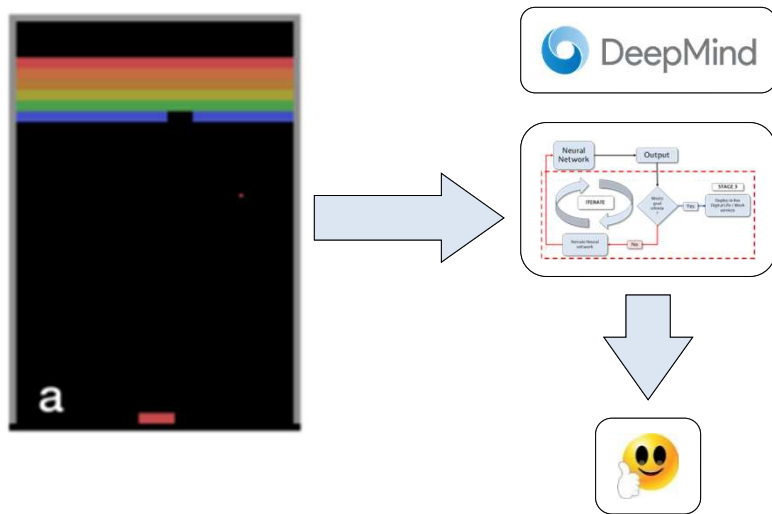
Kansky et al's testing of Deep Mind's algorithm with a changed layout
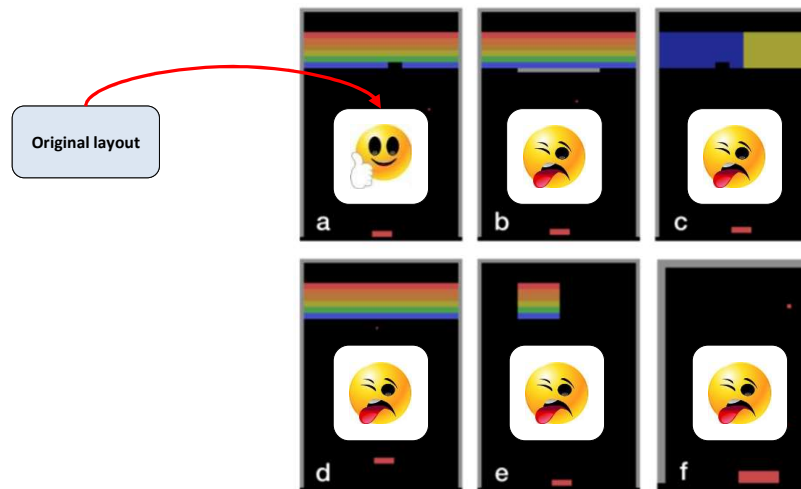
The algorithm is unable to adapt to changes in the game layout demonstrating that there is no causal understanding but merely a statistical relationship between paddle movements and the score. **The system is unable to distinguish between a causation and correlation. Human reconquering of the game of Go (probably temporarily) is a good example This is the essence of goal No. 2: transfer learning.**

Source: Kansky et al. (2017). Schema Networks: Zero-shot Transfer with a Generative Causal Model of Intuitive Physics. arXIv, cs.AI.

www.radiofreemobile.com
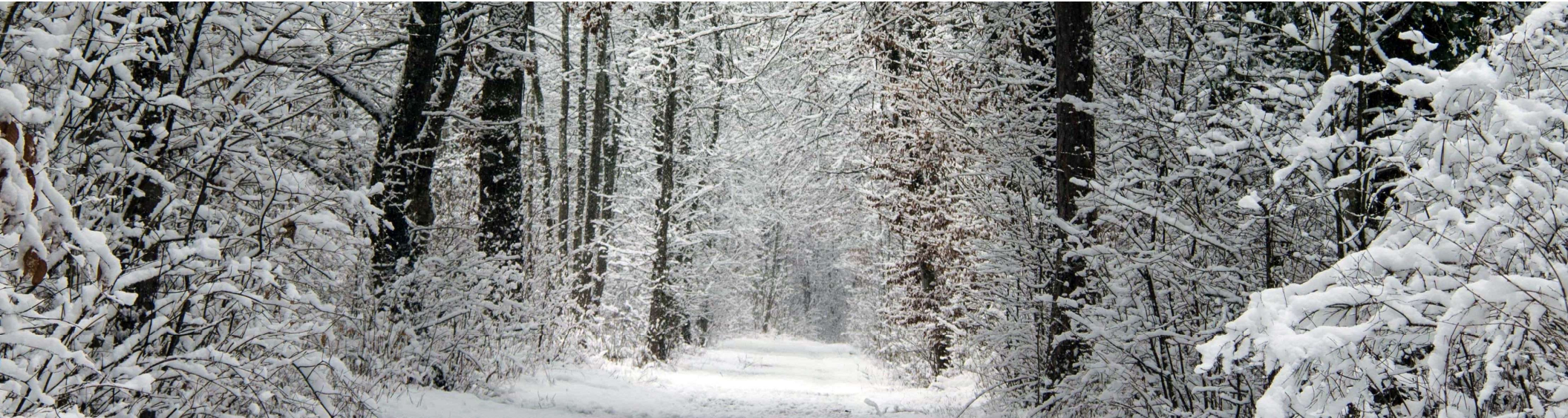
The signals & Economics

Usage, search and business models

ChatGPT should not be seeing declines in traffic…
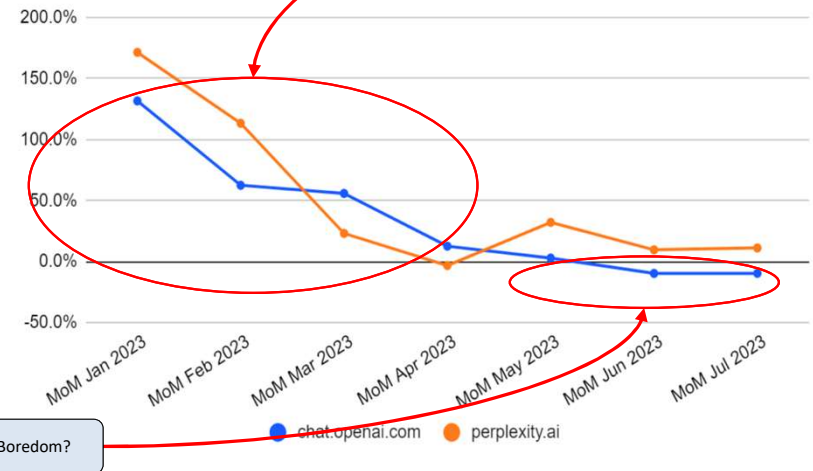
Traffic patterns of ChatGPT vs competitors
January - July 2023, %

Month on Month (MoM) traffic growth by company (June, July 2023)
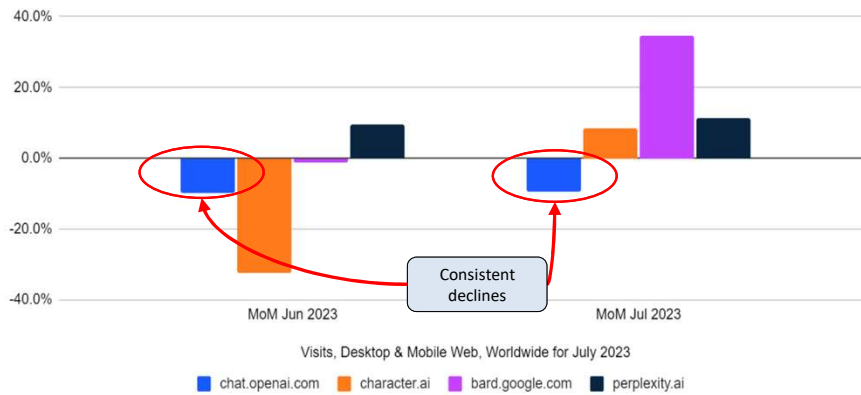
Month on Month traffic growth by company (Jan-Jul 23)

A product that is going to change the world, should not be seeing declines in usage at this stage of its growth. This could easily be a summer-related temporary dip but either way it is not a good sign and points to ChatGPT being craze of which some users are becoming weary.
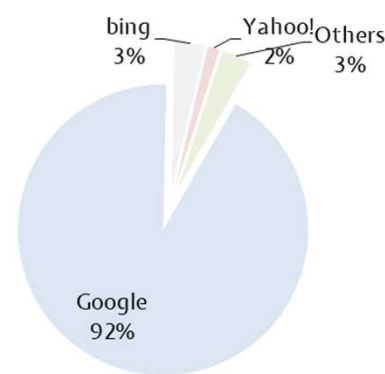
www.radiofreemobile.com

…where it has had no discernible impact on the search industry to date even when it is powering Bing.

Search engine market share
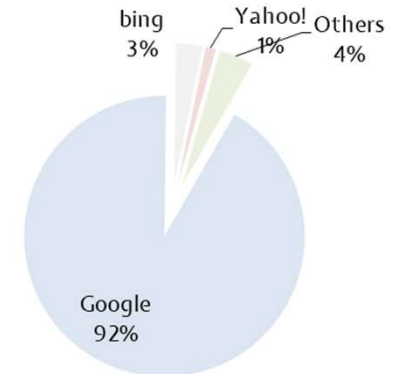*August 2022 – August 2023*

Global search market share desktop & mobile Aug 22 – Aug 23
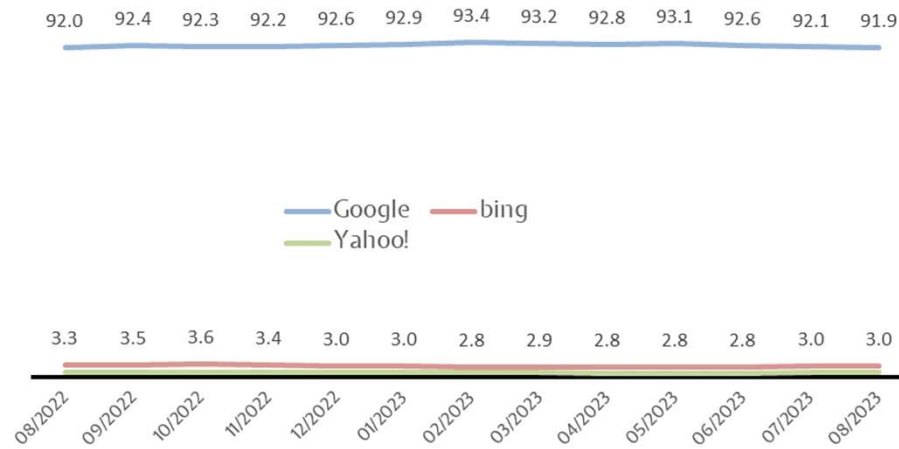
August 2022

August 2023

ChatGPT and generative AI in general has had no appreciable impact on search even though Bing is now powered by GPT-4. If generative AI is replacing search, traffic should not be declining and there should have been some signals in search share data. So far there is none.

www.radiofreemobile.com

The erosion of subscription price for generative AI services is likely to be the first sign of the bubble bursting

Competitive environment for LLM based services
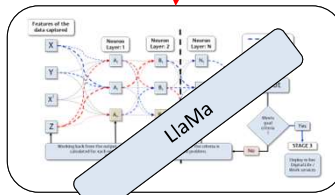*2023*

	Users / price
Microsoft / OpenAI	0.8m / $20
Midjourney	1m / $20
Google Bard	100m / $0
ANTHROP\C	0.1m / $5
Microsoft (Bing)	100m / $0

+

Model weights (W) — LlaMa

LlaMa has enabled hundreds of tuned models to be made available for free
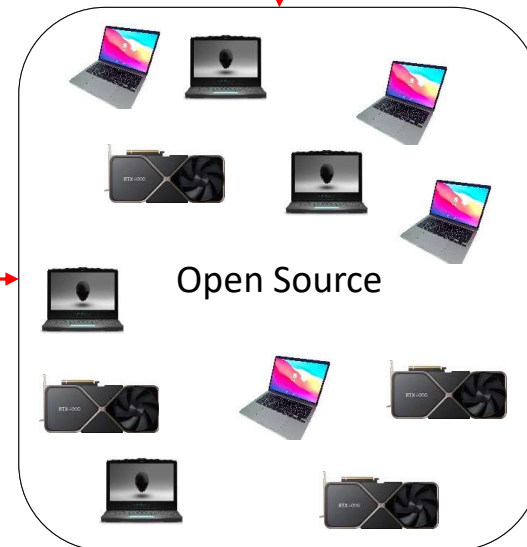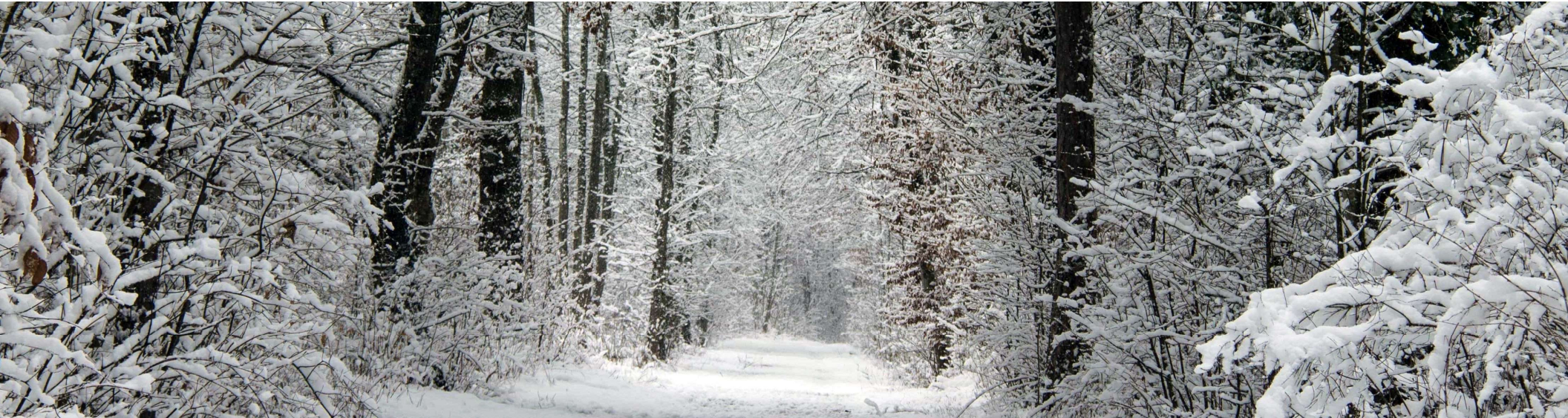
Open Source

With more and more models being fine-tuned and trained by the open-source community, supply is rapidly increasing.

This will lead to increasing pressure on the price of subscription driven generative AI services.

RFM forecasts that a price of $20 per year is far more realistic in the long-term as opposed to the current standard of $20 per month.

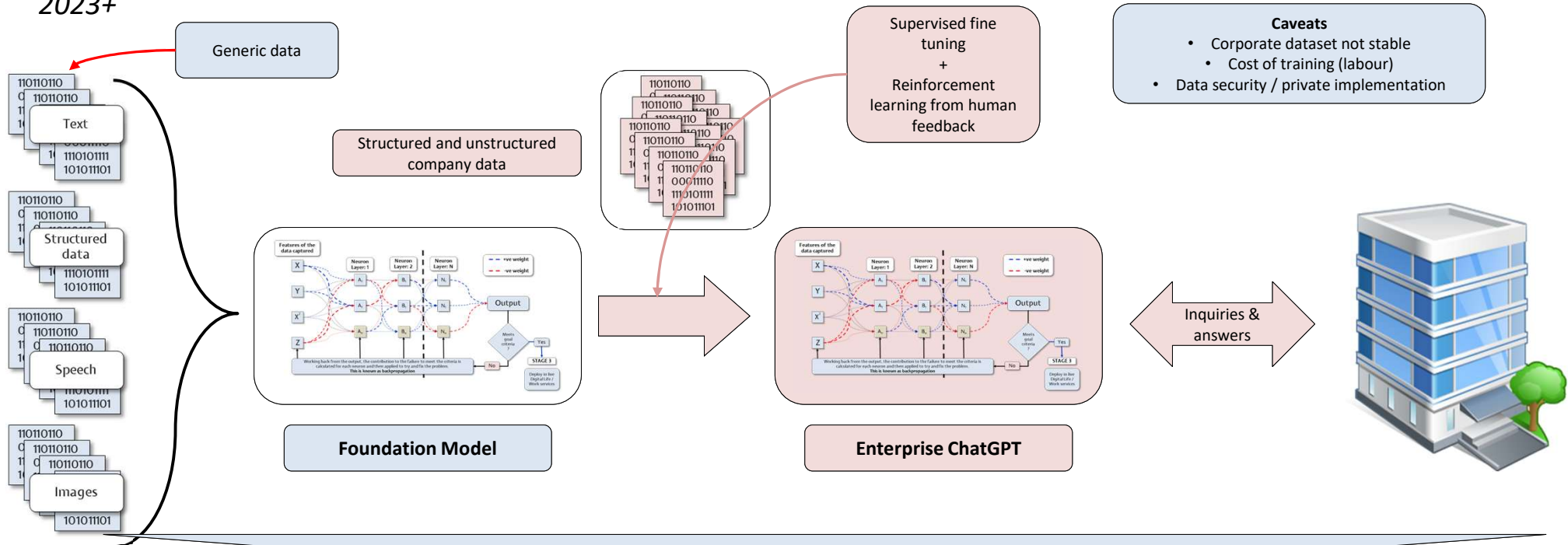**www.radiofreemobile.com**

Use cases of Generative AI

Beyond entertainment

Generative AI offers the possibility to obviate the use of a corporate database

Use case 1: Generative AI in the enterprise
2023+

Generic data

Text

Structured data

Speech

Images

Structured and unstructured company data

Supervised fine tuning
+
Reinforcement learning from human feedback

Caveats
- Corporate dataset not stable
- Cost of training (labour)
- Data security / private implementation

Foundation Model

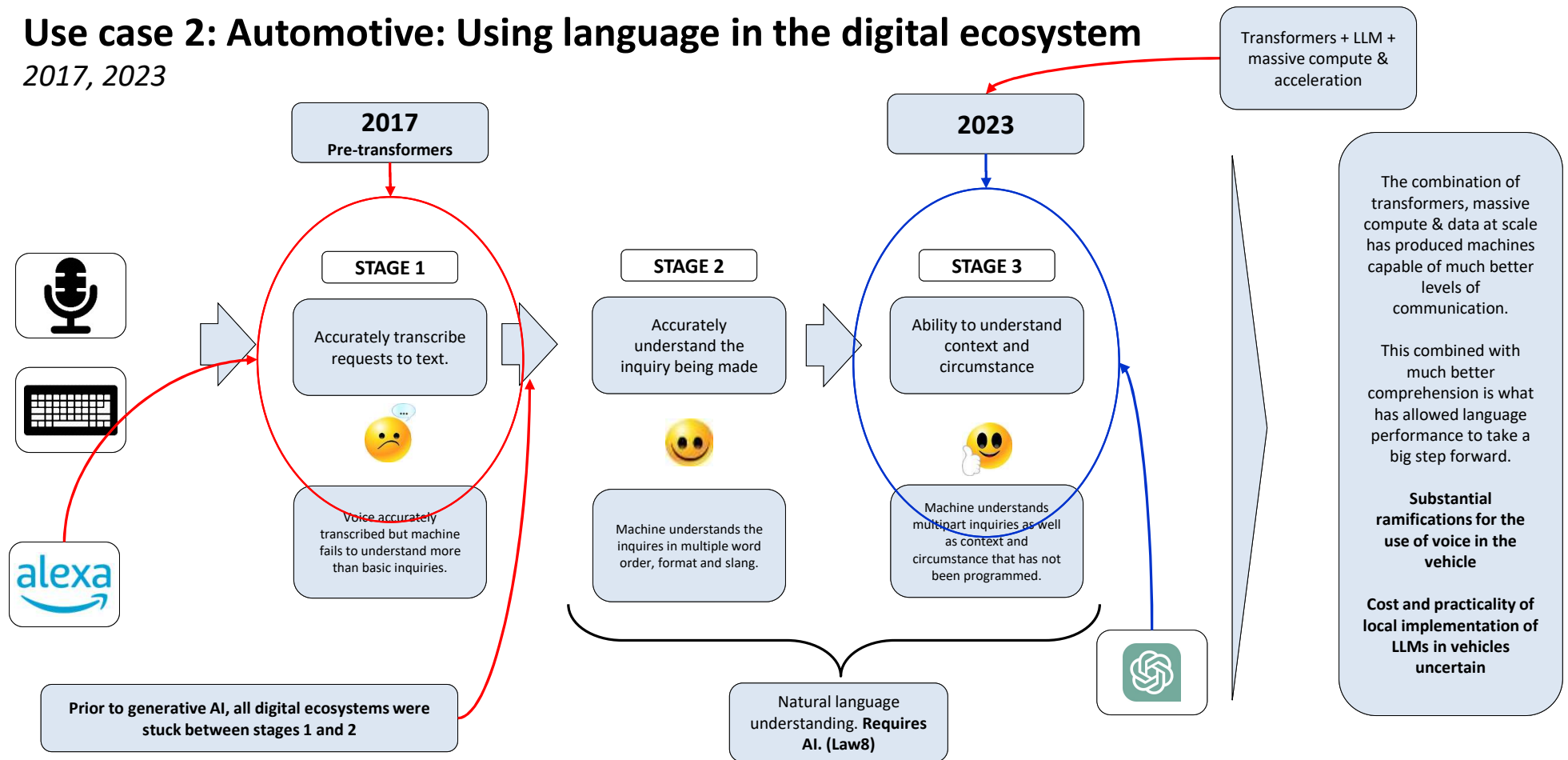**Enterprise ChatGPT**

Inquiries & answers

The need to structure data or use a database largely removed as data can be recalled through language requests. Answers will need to be verified. Including verification, RFM thinks this will result in a substantial cost saving or a productivity enhancement. To deal with caveats, may only be of use for historical data and would need periodic updating.

Source: Radio Free Mobile   **www.radiofreemobile.com**

Transformers and compute have allowed a great improvement in language interaction with machines. Perhaps enough to be used in the vehicle

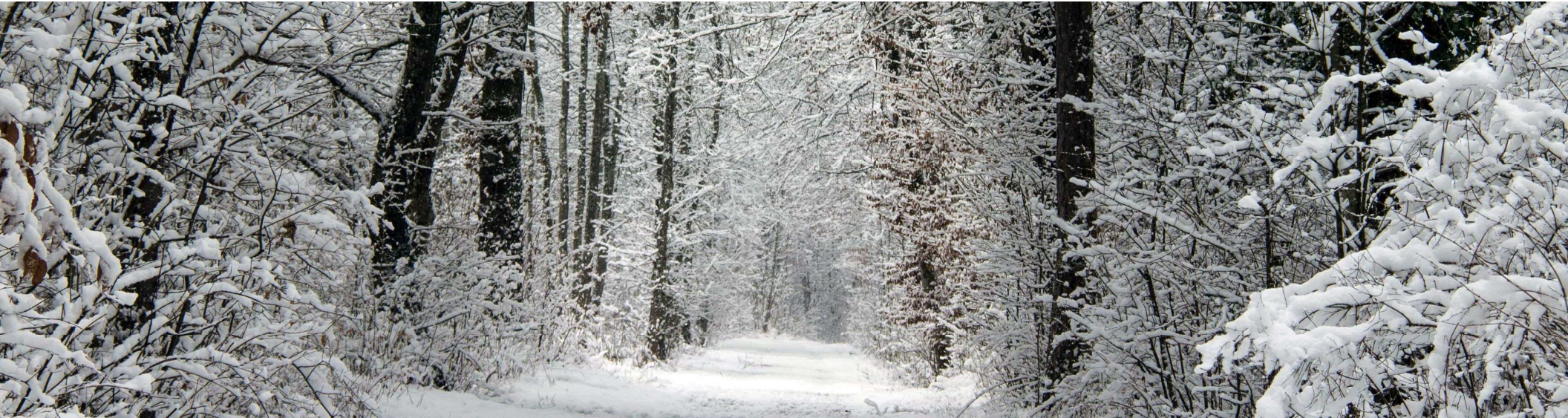# Use case 2: Automotive: Using language in the digital ecosystem
2017, 2023

Transformers + LLM + massive compute & acceleration

2017
Pre-transformers

2023

STAGE 1

Accurately transcribe requests to text.

Voice accurately transcribed but machine fails to understand more than basic inquiries.

STAGE 2

Accurately understand the inquiry being made

Machine understands the inquires in multiple word order, format and slang.

STAGE 3

Ability to understand context and circumstance

Machine understands multipart inquiries as well as context and circumstance that has not been programmed.

The combination of transformers, massive compute & data at scale has produced machines capable of much better levels of communication.

This combined with much better comprehension is what has allowed language performance to take a big step forward.

Substantial ramifications for the use of voice in the vehicle

Cost and practicality of local implementation of LLMs in vehicles uncertain

Prior to generative AI, all digital ecosystems were stuck between stages 1 and 2

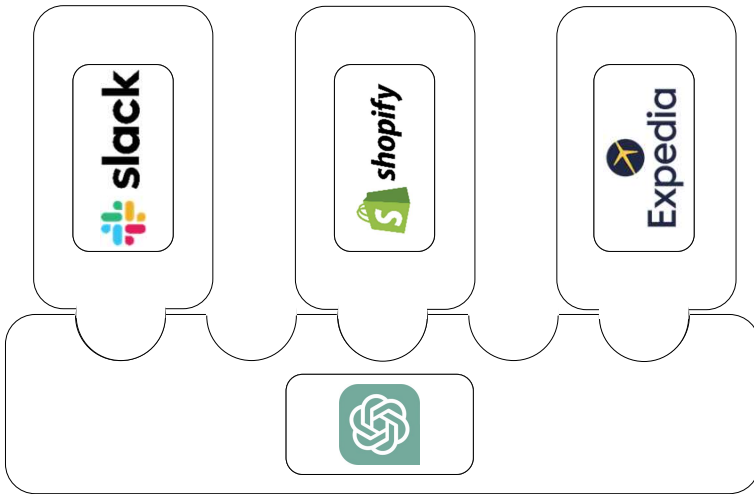Natural language understanding. **Requires AI. (Law8)**

Moats & The Ecosystem

Both OpenAI and Meta are well on the way to having a large moat to defend against competition. Ironically Meta's moat is the open source itself
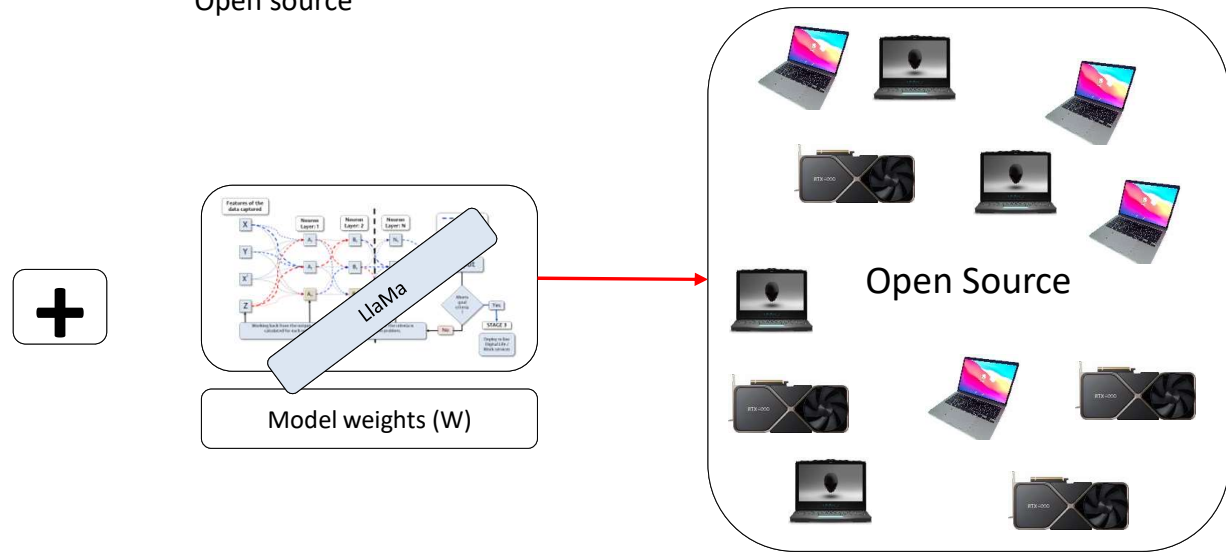
The power of the platform

ChatGPT plugins

Open source

Open Source

Model weights (W)

ChatGPT is becoming a platform for development in its own rights. Its 1bn users provides a great incentive for developers to develop for ChatGPT first. **140+ plugins already created**

Every open source model is based on LlaMa. Foundation models are very hard to switch. Meta is building itself a large moat

Sign up for the daily at www.radiofreemobile.com

To entertain as well as inform

www.radiofreemobile.com