Three challenges with AI where *BitBrain* can help

Michael Hopkins & Jakub Fil
The University of Manchester

30th – 31st May 2023

michael.hopkins@manchester.ac.uk
jakub.fil@manchester.ac.uk

Human Brain Project

MANCHESTER 1824 The University of Manchester

If we ignore some societal issues…

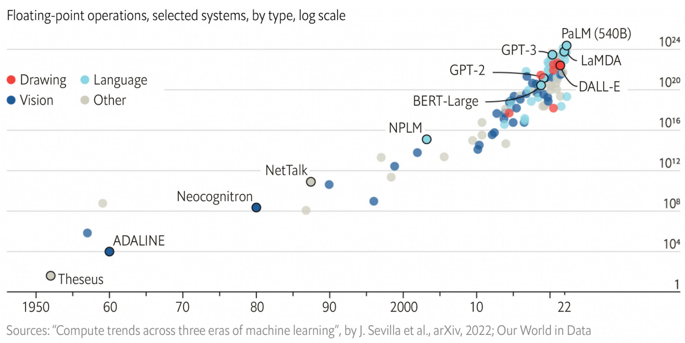
- **Monopoly control of information and choice**
- **Interference in democracy and science**
- **Autonomous weapons**
- **Mass unemployment in unforeseen sectors**
- **Effects of training bias and misinformation**

…and consider only technical aspects…

1 - Unsustainable energy use

- Energy (and CO_2) needed to train latest AI models already huge. Power stations built near the server farms.

1 - Unsustainable energy use

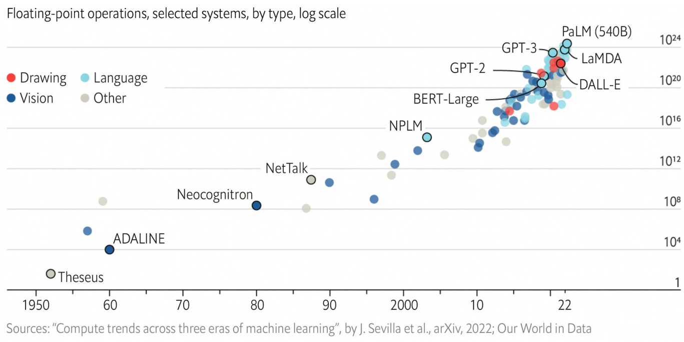


- Energy (and CO_2) needed to train latest AI models already huge. Power stations built near the server farms.
- Cannot continue growing super-exponentially or planet will be covered with memory chips with no electricity left for anything else!

Environmental Impact of Select Machine Learning Models, 2022
Source: Luccioni et al., 2022 | Table: 2023 AI Index Report

Model	Number of Parameters	Grid Carbon Intensity	Power Consumption	C02 Equivalent Emissions
GPT-3	175B	429 gC02eq/kWh	1,287 MWh	502 tonnes

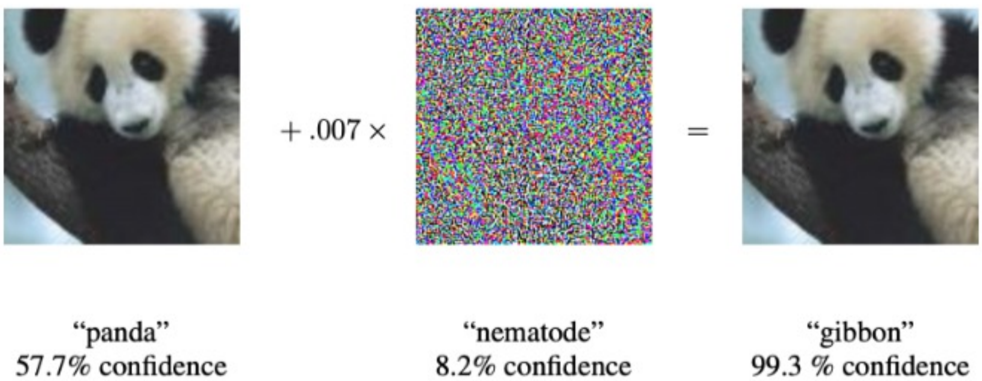
1 - Unsustainable energy use



- **Energy (and CO_2) needed to train latest AI models already huge. Power stations built near the server farms.**
- **Cannot continue growing super-exponentially or planet will be covered with memory chips with no electricity left for anything else!**
- **Autonomous vehicles require AI onboard using ≈1W instead of current ≈100W. IoT/edge devices need to be far more economical still.**
- **Green legislation will make this waste unviable not just very expensive (e.g. training LLaMa = $4M). GPT-4 is ≈1000x bigger than LLaMa…**

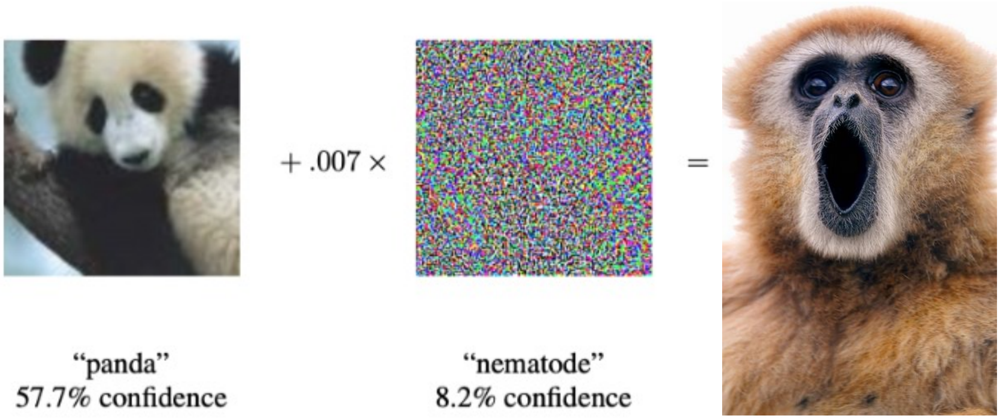
2 – Brittle and hard to explain

- **Vast parameter sets (100 trillion+ for GPT-4) used for current AI models 'over-fit' the data.**
- **Tiny changes (e.g. to an image) can fool the system and make classification incorrect – with very high confidence.**



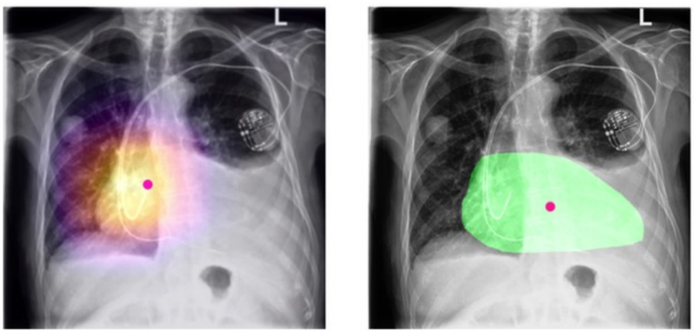
2 – Brittle and hard to explain

- **Vast parameter sets (100 trillion+ for GPT-4) used for current AI models 'over-fit' the data.**
- **Tiny changes (e.g. to an image) can fool the system and make classification incorrect – with very high confidence.**



2 – Brittle and hard to explain

- **Vast parameter sets (100 trillion+ for GPT-4) used for current AI models 'over-fit' the data.**
- **Tiny changes (e.g. to an image) can fool the system and make classification incorrect – with very high confidence.**
- **These changes could be from errors, noise or even malign actors.**
- **The ability to explain and justify results is increasingly a legal requirement.**

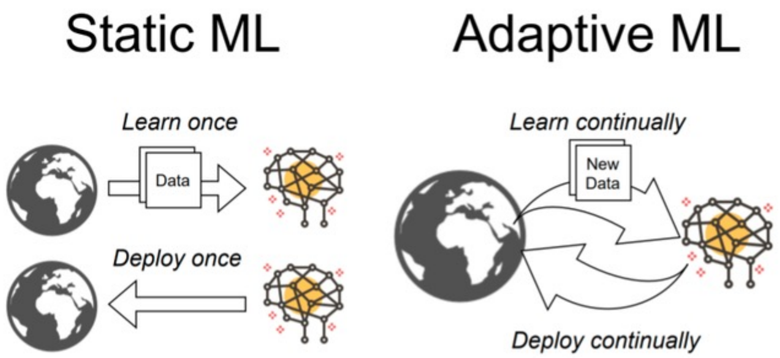


Grad-CAM heat map for Cardiomegaly
- Maximally activated pixel in heat map

Benchmark segmentation for Cardiomegaly
- Single point used to locate pathology by radiologists

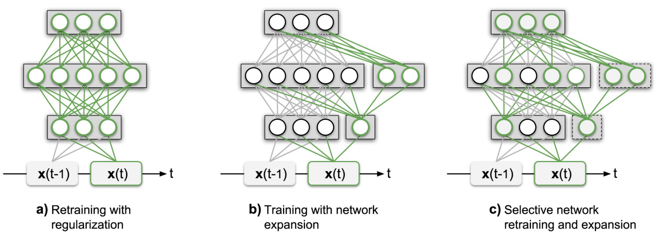
3 - Static and inflexible

- Current AI is trained once on fixed training set at enormous cost, and the learning is frozen.
- Real world is not like that. Everything is constantly changing both within and outside the system.

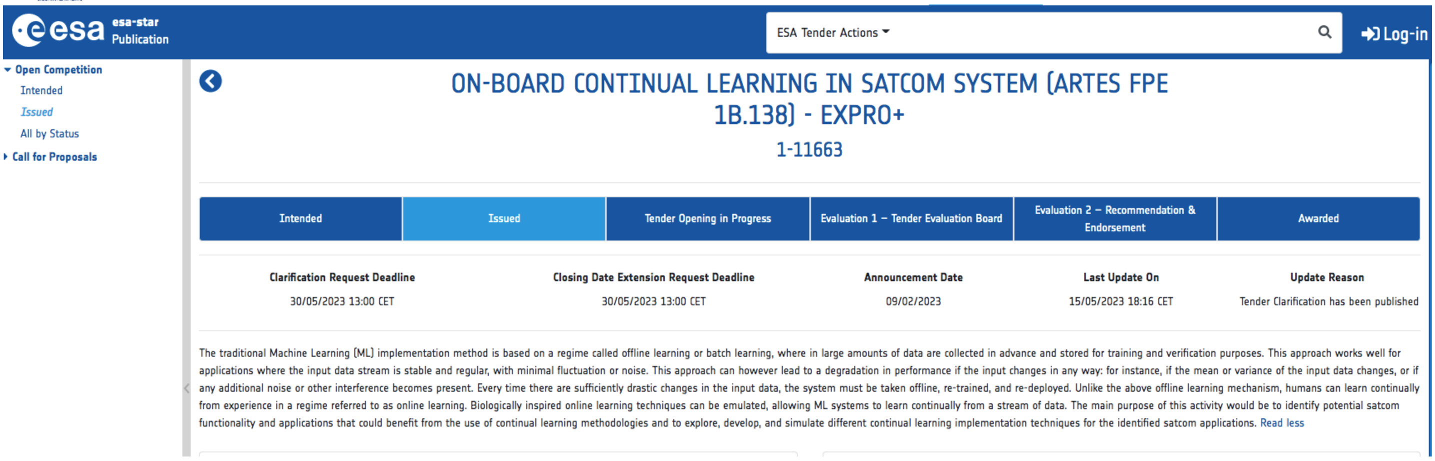


3 - Static and inflexible

- Current AI is trained once on fixed training set at enormous cost, and the learning is frozen.
- Real world is not like that. Everything is constantly changing both within and outside the system.
- Very difficult task to give the necessary flexibility to AI models without ongoing large costs.



3 - Static and inflexible


esa esa-star Publication

ESA Tender Actions

Log-in

- Open Competition
  - Intended
  - Issued
  - All by Status
- Call for Proposals

ON-BOARD CONTINUAL LEARNING IN SATCOM SYSTEM (ARTES FPE 1B.138) - EXPRO+

1-11663

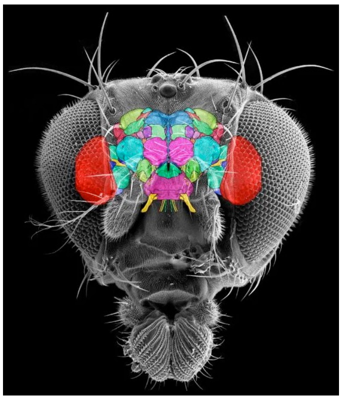
Intended	Issued	Tender Opening in Progress	Evaluation 1 – Tender Evaluation Board	Evaluation 2 – Recommendation & Endorsement	Awarded

Clarification Request Deadline	Closing Date Extension Request Deadline	Announcement Date	Last Update On	Update Reason
30/05/2023 13:00 CET	30/05/2023 13:00 CET	09/02/2023	15/05/2023 18:16 CET	Tender Clarification has been published

The traditional Machine Learning (ML) implementation method is based on a regime called offline learning or batch learning, where in large amounts of data are collected in advance and stored for training and verification purposes. This approach works well for applications where the input data stream is stable and regular, with minimal fluctuation or noise. This approach can however lead to a degradation in performance if the input changes in any way: for instance, if the mean or variance of the input data changes, or if any additional noise or other interference becomes present. Every time there are sufficiently drastic changes in the input data, the system must be taken offline, re-trained, and re-deployed. Unlike the above offline learning mechanism, humans can learn continually from experience in a regime referred to as online learning. Biologically inspired online learning techniques can be emulated, allowing ML systems to learn continually from a stream of data. The main purpose of this activity would be to identify potential satcom functionality and applications that could benefit from the use of continual learning methodologies and to explore, develop, and simulate different continual learning implementation techniques for the identified satcom applications. Read less

- **Hence the growing number of research projects setting out to address this issue.**

Biological systems do things differently e.g.



- **Learn continuously & adapt to changes**
- **Use tiny amounts of energy**
- **Performance degrades gracefully & safely**

...and so we chose to learn from them and create *BitBrain* to address these concerns

Patent GB 2113341.8 filed at the UKIPO in Sept 2021 by MH and Steve Furber

What is *BitBrain*?

Algorithm with unique benefits, especially when combined with hardware

- **Single-pass learning avoids expensive computations**
- **Accurate inference that is naturally robust against poor inputs & uncertainty**
- **Continuous and adaptive learning**
- **Fast and low-energy operation on conventional & neuromorphic processors**
- **Implemented on SpiNNaker, the world's largest real-time brain simulator**

Hopkins, M., Fil, J., Jones, E. G. & Furber, S. (2023); ***BitBrain* and Sparse Binary Coincidence (SBC) memories: Fast, robust learning and inference for neuromorphic architectures**. *Frontiers in Neuroinformatics*

Potential *BitBrain* applications

- **Intelligent medical devices where privacy and low energy use are paramount**
- **IoT/Edge computing intelligent sensors**
- **High-throughput security screening on e.g. network switches**
- **Adaptive robotics making robust decisions in changing environments**
- **Efficient on-chip processing for current and future sensors e.g. event-based**
- **Situations where custom models need to be learned for every individual device in a high-volume manufacturing context**
- **Leveraging neuromorphic and future hardware (e.g. memristors) to exploit speed and energy efficiency**
- **Applicable to many data types in a multi-modal fashion e.g. images, audio, codes, sequences, time series, graphs, …**

Current status and outlook

Now

***BitBrain* works well on classic data sets of various kinds and results are published.**

Next

Theory development and application to more complex data sets and tasks.
...which requires **Expansion of the team** – both technical and management.

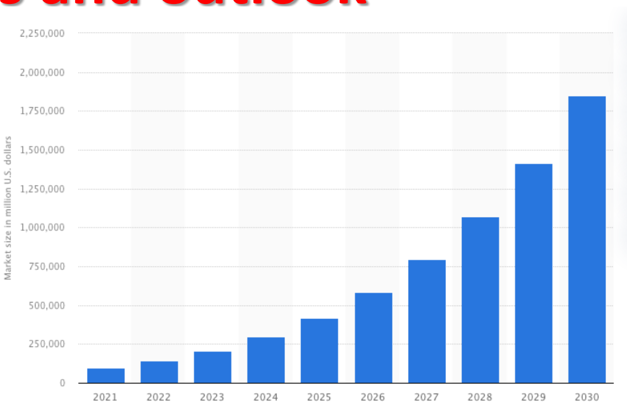
Current status and outlook

Now

***BitBrain* works well on classic data sets**

Next

Theory development and application to
...which requires **Expansion of the team**



Addressable market size

Global AI market in 2022 ≈$136 billion. Projected growth to 2030 ≈37% per annum.
Global IoT market in 2022 ≈$544 billion. Projected growth to 2030 ≈26% per annum.

Any market estimate is a guess, but any realistic fraction will be substantial.

Current status and outlook

Now

***BitBrain* works well on classic data sets of various kinds and results are published.**

Next

Theory development and application to more complex data sets and tasks.
...which requires **Expansion of the team** – both technical and management.

Addressable market size

Global AI market in 2022 ≈$136 billion. Projected growth to 2030 ≈37% per annum.
Global IoT market in 2022 ≈$544 billion. Projected growth to 2030 ≈26% per annum.

Any market estimate is a guess, but any realistic fraction will be substantial.

With your help we can take the next big step in AI

Selected *BitBrain* results

Michael Hopkins & Jakub Fil
The University of Manchester

May 2023

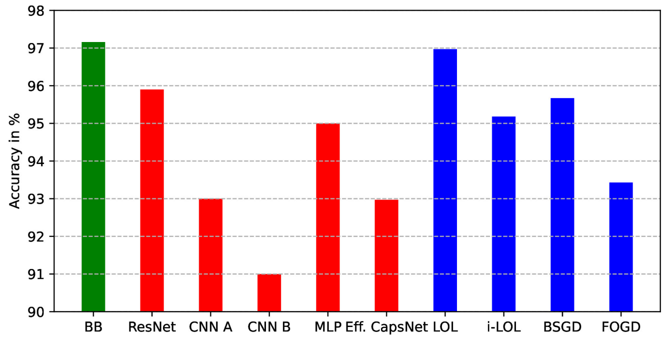
michael.hopkins@manchester.ac.uk

jakub.fil@manchester.ac.uk

Human Brain Project

MANCHESTER 1824
The University of Manchester

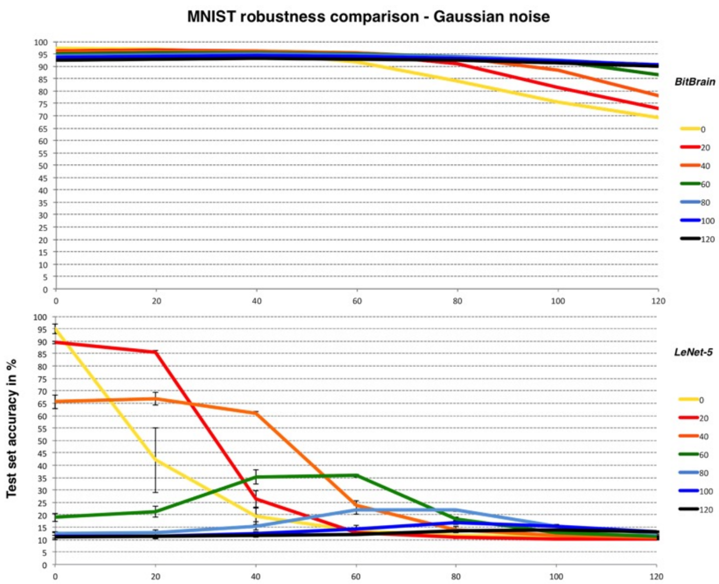
Single-pass and Single-shot learning



Task	libSVM	Perceptron	Pegasos1	Pegasos20	LASVM	StreamSVM1	StreamSVM2	*BitBrain*
0 vs 1	99.52	99.47	95.06	99.48	98.82	99.34	99.71	99.95
8 vs 9	96.57	95.90	69.41	90.62	90.32	84.75	94.70	98.49

- Unlike real neuronal systems, deep learning methods require a huge number of repetitions of the inputs in order to learn.
- *BitBrain* is a state-of-the-art single-pass method - it only sees the data once and therefore learns very quickly and uses far less energy.
- *BitBrain* shows excellent performance on single-shot learning tasks, where the model is given only one example of a new type of input.
- This combination offers continuous and adaptive learning, which is usually either difficult or infeasible with deep neural networks.

Robust against imperfect inputs



- *BitBrain* performs much better than many deep neural networks given imperfect inputs.
- It can tolerate large amounts of noise and other problems in the data required for making decisions. This makes if far less 'brittle' than other methods.
- This graceful degradation like real neural systems is a very useful feature, especially when the system is deployed, e.g. in a noisy environment, with changing conditions or sensor degradation.

BitBrain Business Canvas and Budget Overview

Michael Hopkins & Jakub Fil
The University of Manchester

May 2023

michael.hopkins@manchester.ac.uk

jakub.fil@manchester.ac.uk

Human Brain Project

MANCHESTER 1824
The University of Manchester

Value proposition

We aim to address technical shortcomings of modern AI models, whilst at the same time making AI greener and more sustainable. We offer a unique technology which addresses many of the market needs:

- **Single-pass and Single-shot learning**
- **Noise and adversarial resilience**
- **Adaptive and continuous learning**
- **Energy efficient learning and decision making**
- **Perfect match to IoT paradigm**

Key Activities and Channels

Key Activities:

- **Further development of the theory and practical applications - benchmarking against industrial-grade datasets**
- **Building an interface which may be used by the third parties**
- **Exploring new types of hardware for even faster learning and lower energy use – FPGAs, memristors, event-based sensors, 3D memory, neuromorphic chips**

Channels:

- **Individual licensing of the technology**
- **Delivering individually fitted solutions to industrial needs of our partners**
- **Online self-service application and an easily accessible API**

Key partners and Resources

Key partners:

- **Public institutions e.g. ESA or NGOs.**
- **Industrial partners who require low energy / noise resilient / single-shot / continuous learning AI solutions**
- **Business angels and VCs with a social and/or green agenda**

Key Resources:

- **Patented technology**
- **Experts with a track record in neuromorphic and AI technologies, as well as technology consulting**
- **In-house access to the worlds largest real-time brain simulator (SpiNNaker)**
- **Prof Steve Furber available on the board in an advisory capacity**

Customer relationships and Revenue streams

Customer relationships :

- **Individual and on demand solutions for specific problems**
- **Personal assistance in understanding the needs of our partners**
- **Online self-service application based on either standard server-grade processors and the SpiNNaker cluster in Manchester.**

Revenue streams:

In order of preference

1. **Licensing of IP with specific consultancy appropriate for each end-user/task.**
2. **Develop proprietary solutions in conjunction with collaborators.**
3. **Develop specific H/W with collaborators.**

Customer segments

Markets which require specific technology solutions

- **Industrial applications which require continuous and adaptive learning**
- **Personal medical devices where privacy and low energy use are paramount**
- **IoT/Edge computing intelligent sensors e.g., the ML sensors 2.0 paradigm**
- **High-throughput security screening on e.g., network switches**
- **Adaptive robotics making robust decisions in changing environments**
- **Quick, intelligent, and low-energy receivers for event-based sensors**
- **Situations where custom models need to be learned for every individual device in a high-volume manufacturing context**
- **Leveraging neuromorphic & future H/W (e.g. memristors) to maximise speed & energy efficiency**
- **Data fusion for many data types e.g. images, audio, codes, sequences, time series, graphs, …**

Cost Structure

Fixed costs

Per annum £720k office costs and employment of technical and business staff

- **4 technical people**
- **2 business people (some part time)**

Variable costs

- **Research costs - development of new features, securing necessary computation resources to benchmark against other solutions**
- **Exploring new types of hardware which may enable our technology to further improve its characteristics - FPGAs, memristors, future neuromorphic processors, event-based sensors**

Budget overview

Staff

4x technical staff (development of theory, algorithms and potentially hardware)

2x business staff (legal, marketing, business structure)

Period

1-2 years is an estimate of time needed to show results suitable for commercial exploitation.

Costs

£120k per staff member per year, inclusive of all overheads (some staff may be part-time).

Other sources of income

Academic research proposals in the pipeline (ESA, NWPST, Internal UoM memristor wafer).

Talking to other business angels through personal contacts.

Business model (in order of preference)

Licensing of IP with specific consultancy appropriate for each end-user/task.

Develop proprietary solutions in conjunction with collaborators.