International Railway
Safety Council Conference

Perth I 16 October 2019

Digital Train Control

Functional Safety for AI based Systems

GHD Transportation

David Milburn & Mike Erskine

Grades of Automation

Adapted from **IEC 62290-1** Urban guided transport (UGT) management and command/control systems and UITP World Report on Automation

IEC

IEC 62290-1

INTERNATIONAL STANDARD

NORME INTERNATIONALE

Increasing Automation

		Basic Functions			
		Driving	Supervise Guideway	Supervise Passenger Transfer	Operation during disruption
GoA1	ATP with Driver	Driver	Driver	Driver/Guard /Platform Staff	Driver /Guard
GoA2	ATP and ATO with Driver	Automatic	Driver	Driver/Guard /Platform Staff	Driver /Guard
GoA3	Driverless (DTO)	Automatic	Automatic	Train attendant	Train attendant
GoA4	Unattended (UTO)	Automatic	Automatic	Automatic	Automatic and/or OCC Staff

Grades of Automation

Adapted from **IEC 62290-1** Urban guided transport (UGT) management and command/control systems and UITP World Report on Automation

Increasing Automation

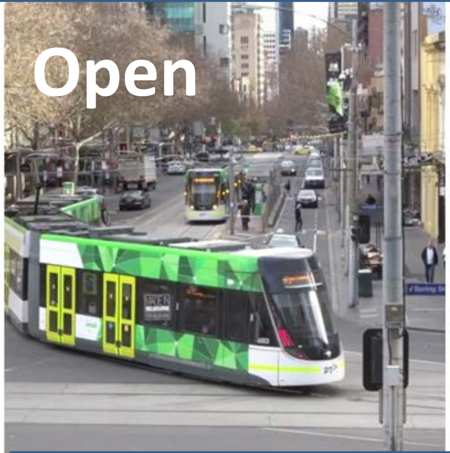
		Basic Functions: Driving	Supervise Guideway	Supervise Passenger Transfer	Operation during disruption
GoA1	ATP with Driver	Driver	Driver	Driver/Guard /Platform Staff	Driver /Guard
GoA2	ATP and ATO with Driver	Automatic	Driver	Driver/Guard /Platform Staff	Driver /Guard
GoA3	Driverless (DTO)	Automatic	Automatic	Train attendant	Train attendant
GoA4	Unattended (UTO)	Automatic	Automatic	Automatic	Automatic and/or OCC Staff
Sydney Metro, CBTC (GoA4)	Unattended (UTO)	Automatic	Automatic? (Alarms)	Automatic (PSD)	OCC Staff
Cross River Rail, ETCS (GoA?)	ATP and ATO with Driver	Automatic	Driver	Automatic (PSD)	Driver
Rio Tinto, ETCS (GoA?)	Unattended (UTO)	Automatic	Level crossings only?	n/a	OCC Staff

The Future - Automatic versus Autonomous

Automatic System: performs task sequences based on pre-defined rules. The information required to understand the environment is provided to enable the system to undertake rehearsed actions. ***Predominantly Deterministic***.

DTO/UTO achieved by Automatic Systems

Autonomous System: capable of making independent decisions to respond to all cases in real-time, and in some situations without reference to pre-defined instructions. It must therefore manage the functions of comprehension, environmental awareness, and spontaneous decision making.
Predominantly Non Deterministic.



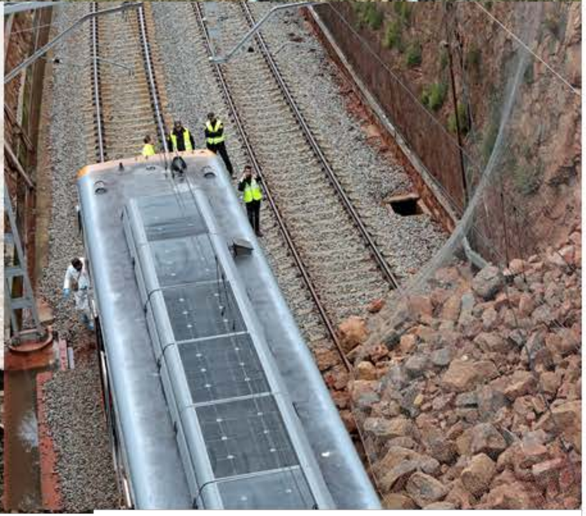
DTO/UTO will be delivered by Autonomous Systems?

Supervise guideway

Supervise Guideway


Supervise guideway
Supervise Guideway

Supervise guideway

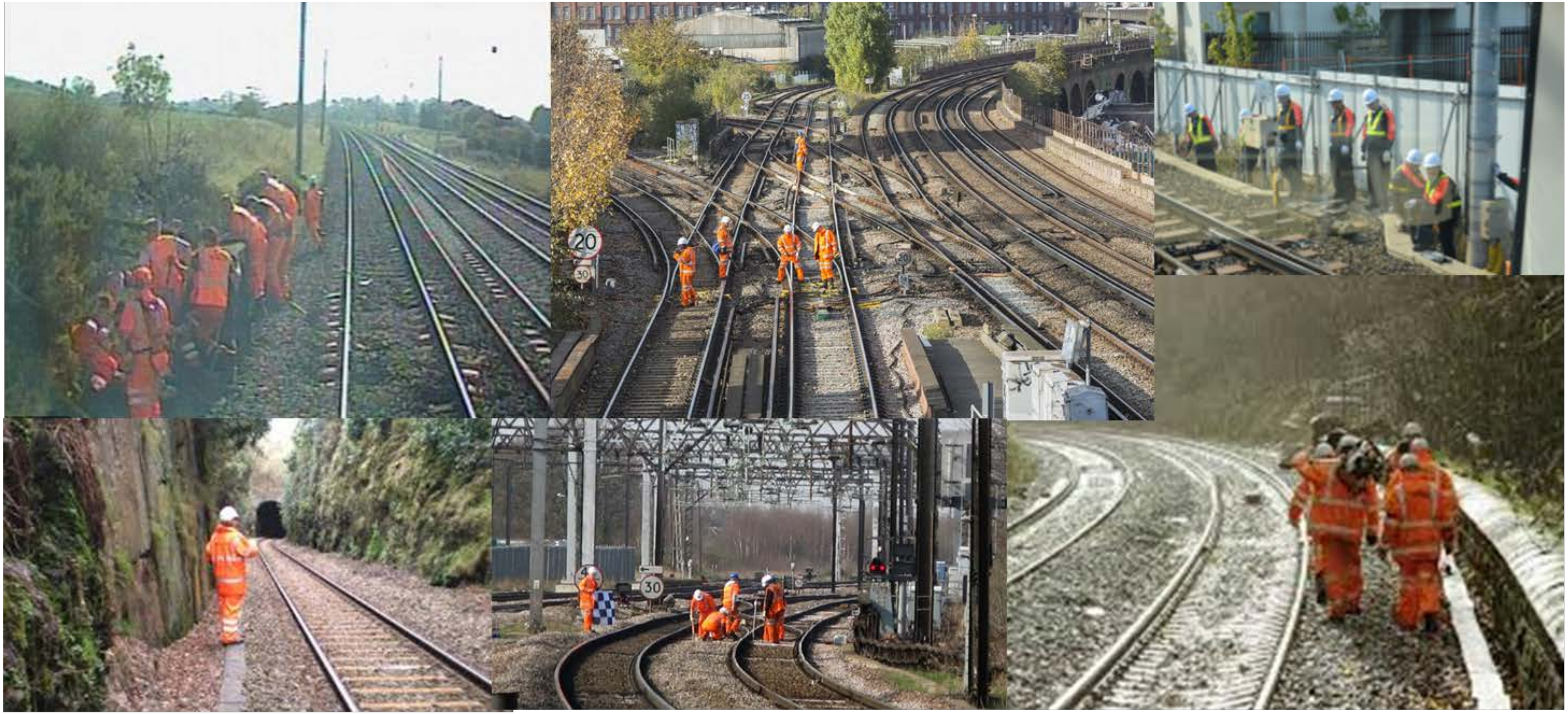


Supervise guideway

Supervise Guideway

Supervise guideway

Supervise Guideway



Supervise guideway

Supervise Guideway

Report infrastructure issues

Report rolling stock issues

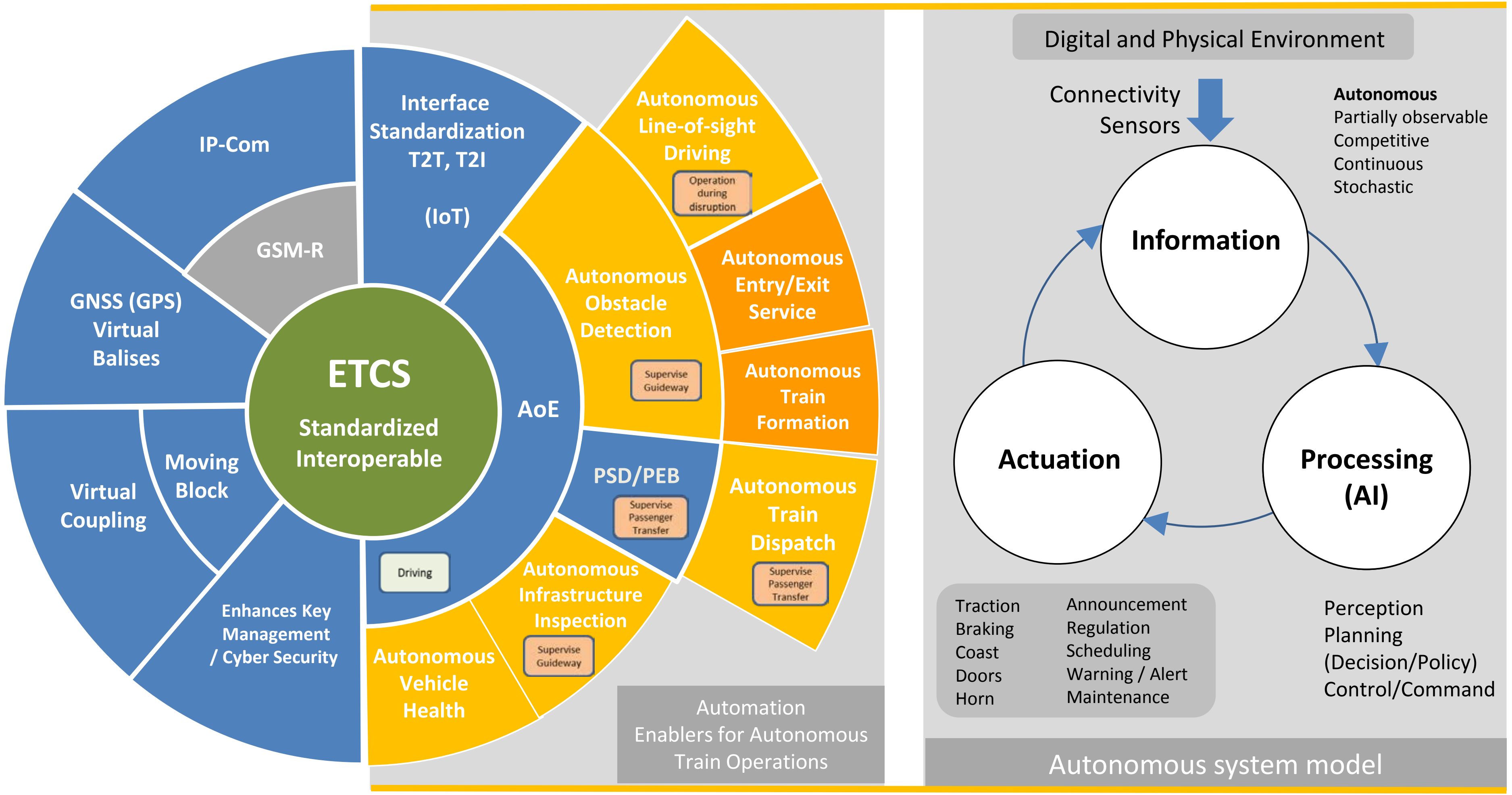




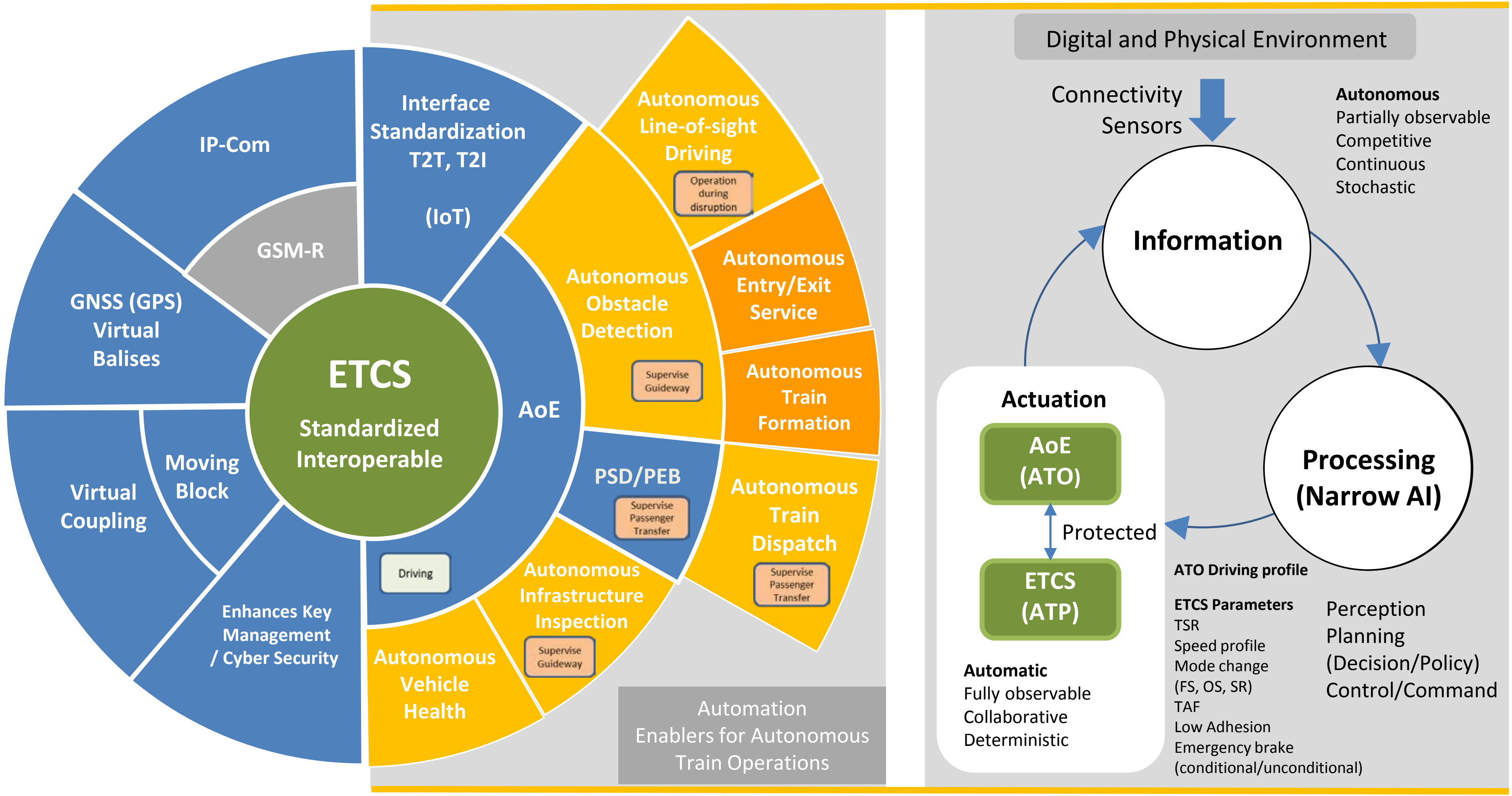

Monitor Vehicle Health

Supervise Passenger Transfer

Building on existing technology - ETCS example (context and evolution)



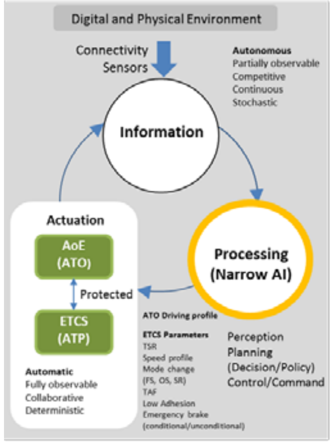
Building on existing technology - ETCS example (context and evolution)

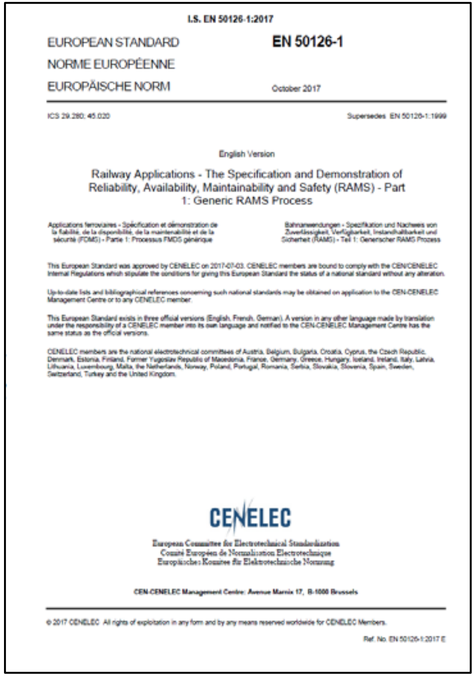


Functional safety for autonomous systems

Autonomous driving systems will use **AI**, this is likely to be based on commercial autonomous car technology (with likely evolution of ISO 26262)

Railway Functional Safety <u>methods</u> will need to <u>adapt</u>





I.S. EN 50126-1:2017

EUROPEAN STANDARD — **EN 50126-1**
NORME EUROPÉENNE
EUROPÄISCHE NORM — October 2017

ICS 29.280; 45.020 — Supersedes EN 50126-1:1999

English Version

Railway Applications - The Specification and Demonstration of Reliability, Availability, Maintainability and Safety (RAMS) - Part 1: Generic RAMS Process

Applications ferroviaires - Spécification et démonstration de la fiabilité, de la disponibilité, de la maintenabilité et de la sécurité (FDMS) - Partie 1: Processus FMDS générique

Bahnanwendungen - Spezifikation und Nachweis von Zuverlässigkeit, Verfügbarkeit, Instandhaltbarkeit und Sicherheit (RAMS) - Teil 1: Generischer RAMS Prozess

This European Standard was approved by CENELEC on 2017-07-03. CENELEC members are bound to comply with the CEN/CENELEC Internal Regulations which stipulate the conditions for giving this European Standard the status of a national standard without any alteration.

Up-to-date lists and bibliographical references concerning such national standards may be obtained on application to the CEN-CENELEC Management Centre or to any CENELEC member.

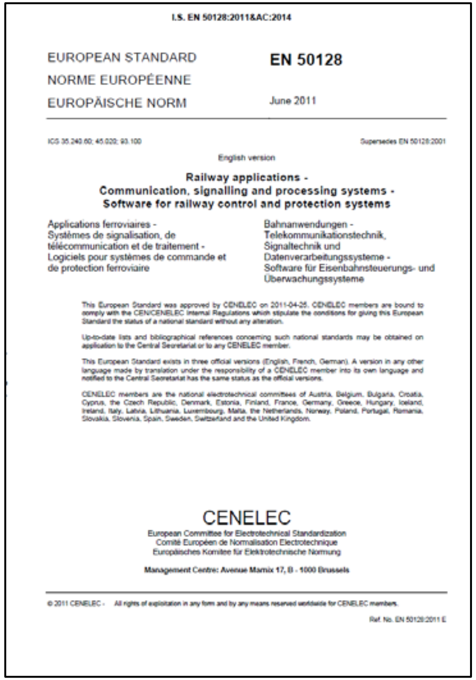
This European Standard exists in three official versions (English, French, German). A version in any other language made by translation under the responsibility of a CENELEC member into its own language and notified to the CEN-CENELEC Management Centre has the same status as the official versions.

CENELEC members are the national electrotechnical committees of Austria, Belgium, Bulgaria, Croatia, Cyprus, the Czech Republic, Denmark, Estonia, Finland, Former Yugoslav Republic of Macedonia, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey and the United Kingdom.

CENELEC
European Committee for Electrotechnical Standardization
Comité Européen de Normalisation Electrotechnique
Europäisches Komitee für Elektrotechnische Normung

CEN-CENELEC Management Centre: Avenue Marnix 17, B-1000 Brussels

Ref. No. EN 50126-1:2017 E



I.S. EN 50128:2011&AC:2014

EUROPEAN STANDARD — **EN 50128**
NORME EUROPÉENNE
EUROPÄISCHE NORM — June 2011

ICS 35.240.60; 45.020; 93.100 — Supersedes EN 50128:2001

English version

Railway applications - Communication, signalling and processing systems - Software for railway control and protection systems

Applications ferroviaires - Systèmes de signalisation, de télécommunication et de traitement - Logiciels pour systèmes de commande et de protection ferroviaire

Bahnanwendungen - Telekommunikationstechnik, Signaltechnik und Datenverarbeitungssysteme - Software für Eisenbahnsteuerungs- und Überwachungssysteme

This European Standard was approved by CENELEC on 2011-04-25. CENELEC members are bound to comply with the CEN/CENELEC Internal Regulations which stipulate the conditions for giving this European Standard the status of a national standard without any alteration.

Up-to-date lists and bibliographical references concerning such national standards may be obtained on application to the Central Secretariat or to any CENELEC member.

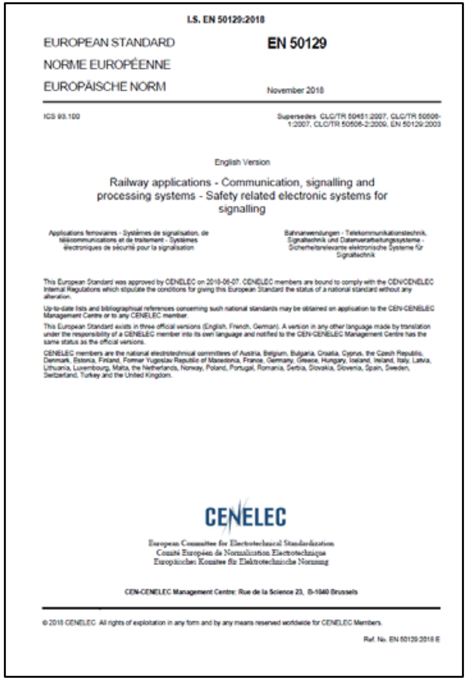
This European Standard exists in three official versions (English, French, German). A version in any other language made by translation under the responsibility of a CENELEC member into its own language and notified to the Central Secretariat has the same status as the official versions.

CENELEC members are the national electrotechnical committees of Austria, Belgium, Bulgaria, Croatia, Cyprus, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland and the United Kingdom.

CENELEC
European Committee for Electrotechnical Standardization
Comité Européen de Normalisation Electrotechnique
Europäisches Komitee für Elektrotechnische Normung

Management Centre: Avenue Marnix 17, B - 1000 Brussels

Ref. No. EN 50128:2011 E



I.S. EN 50129:2018

EUROPEAN STANDARD — **EN 50129**
NORME EUROPÉENNE
EUROPÄISCHE NORM — November 2018

ICS 93.100 — Supersedes CLC/TR 50451:2007, CLC/TR 50506-1:2007, CLC/TR 50506-2:2009, EN 50129:2003

English Version

Railway applications - Communication, signalling and processing systems - Safety related electronic systems for signalling

Applications ferroviaires - Systèmes de signalisation, de télécommunications et de traitement - Systèmes électroniques de sécurité pour la signalisation

Bahnanwendungen - Telekommunikationstechnik, Signaltechnik und Datenverarbeitungssysteme - Sicherheitsrelevante elektronische Systeme für Signaltechnik

This European Standard was approved by CENELEC on 2018-06-07. CENELEC members are bound to comply with the CEN/CENELEC Internal Regulations which stipulate the conditions for giving this European Standard the status of a national standard without any alteration.

Up-to-date lists and bibliographical references concerning such national standards may be obtained on application to the CEN-CENELEC Management Centre or to any CENELEC member.

This European Standard exists in three official versions (English, French, German). A version in any other language made by translation under the responsibility of a CENELEC member into its own language and notified to the CEN-CENELEC Management Centre has the same status as the official versions.

CENELEC members are the national electrotechnical committees of Austria, Belgium, Bulgaria, Croatia, Cyprus, the Czech Republic, Denmark, Estonia, Finland, Former Yugoslav Republic of Macedonia, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey and the United Kingdom.

CENELEC
European Committee for Electrotechnical Standardization
Comité Européen de Normalisation Electrotechnique
Europäisches Komitee für Elektrotechnische Normung

CEN-CENELEC Management Centre: Rue de la Science 23, B-1040 Brussels

Ref. No. EN 50129:2018 E

Intelligence

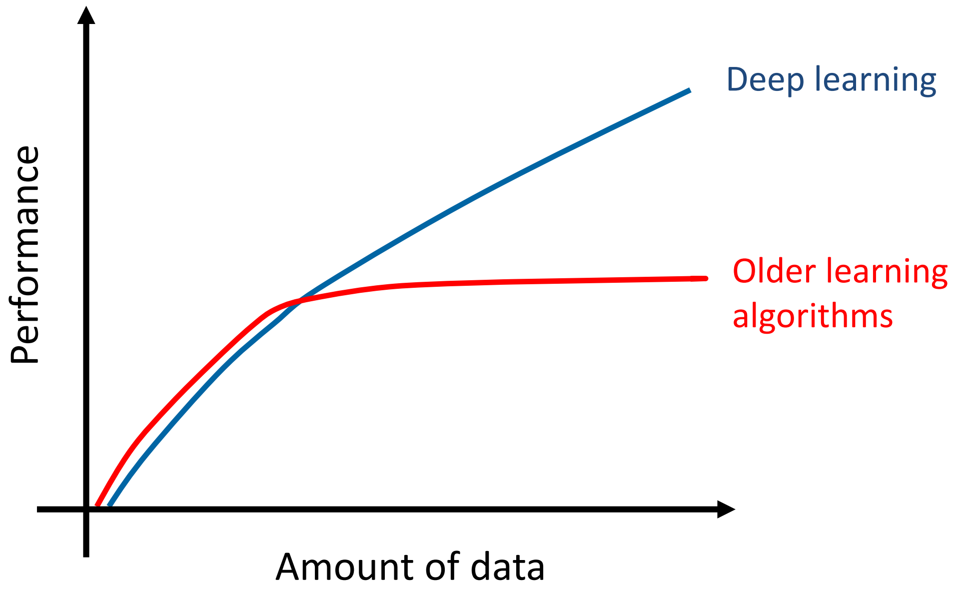
The ability to learn or understand or to deal with new or trying situations.

Artificial Intelligence

Both the intelligence of machines and the branch of computer science which aims to create it, through "the study and design of intelligent agents" or "rational agents", where an intelligent agent is a system that perceives its environment and takes actions which maximize its chances of success.

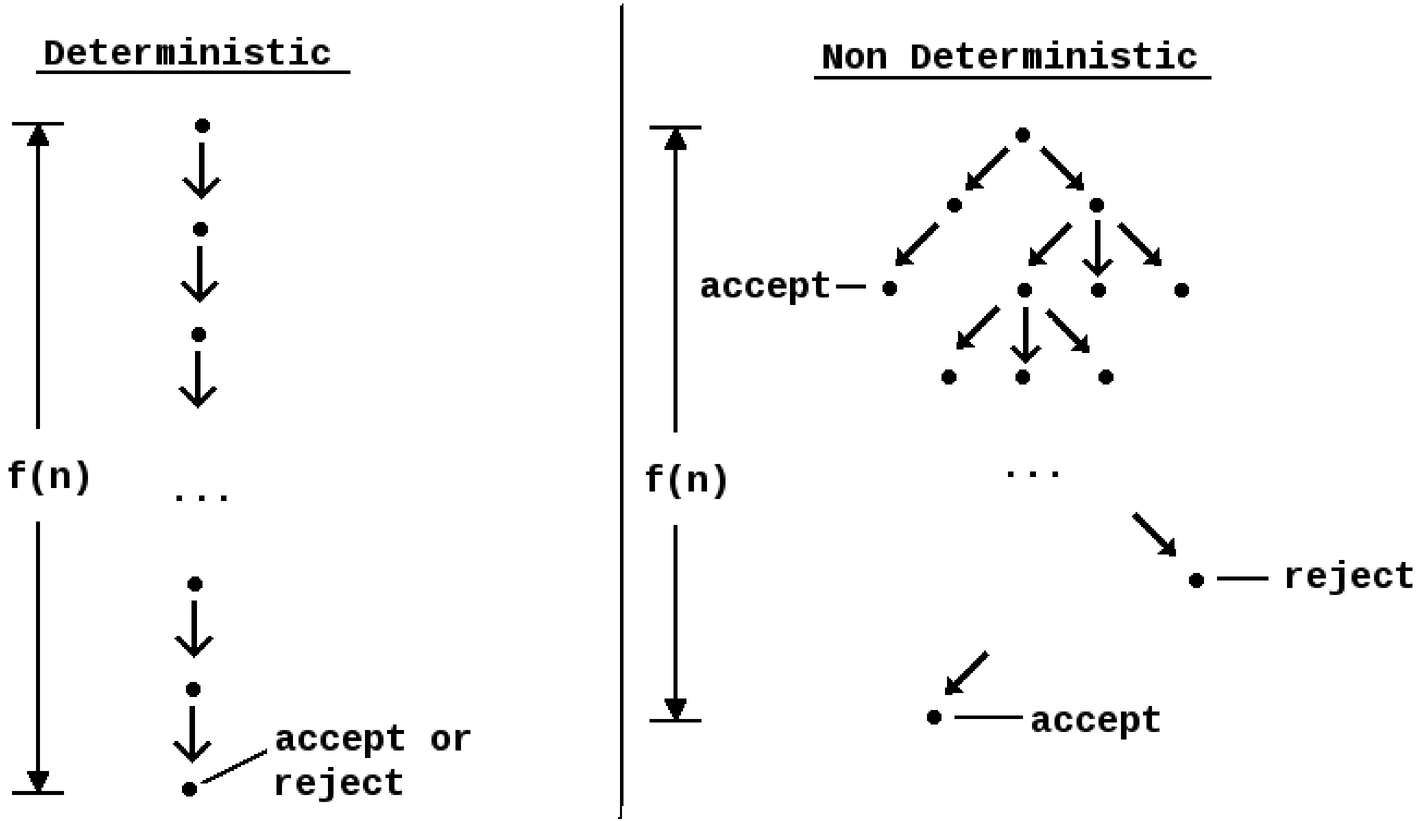
HI + AI = BI ?

Evolution of AI – why deep learning?



How do data science techniques scale with the amount of data?

Deterministic versus non-deterministic

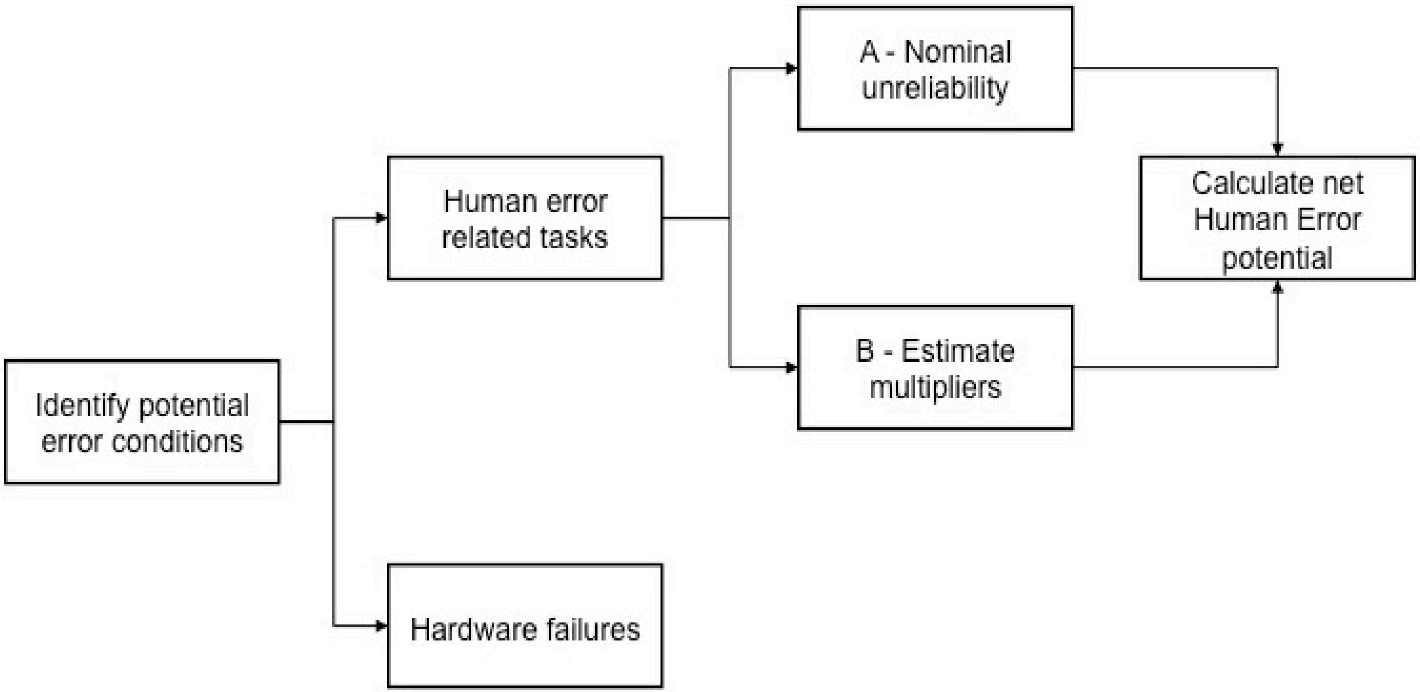


Non-deterministic vs. Stochastic

- Non-deterministic means we know there is uncertainty *but do not know the distribution of uncertainty*
  - *NON-DETERMINISTIC != Uniform distribution*
- E.g. The agent may be in one of {s1,….sn}. We don't know which of them are more vs. less likely (we also don't know that they are all equally likely)
- E.g. the action a done in s can lead to {s1..sm}—no information is available on which outcomes are more likely

- Stochastic is Non-deterministic + distribution information.
- So stochastic means *more knowledge*
  - Note that more knowledge implies more problems are expressible and solvable (but also might mean that the computational burden on the agent increases). With non-determinism, the agent has to say whether there is a strong vs. weak plan. With stochastic actions, the agent can also talk about plans that satisfy goals with different levels of probability

HEART - Human Error Assessment & Reduction Technique



HEART Error Producing Conditions

Generic Task	Nominal Error Probability
Totally unfamiliar, performed at speed with no idea of likely consequences	**0.55** *(0.35 – 0.97)*
Shift or restore system to a new or original state on a single attempt without supervision or procedures.	**0.26** *(0.14 – 0.42)*
Complex task requiring high level of comprehension and skill.	**0.16** *(0.12 – 0.28)*
Fairly simple task performed rapidly or given scant attention.	**0.09** *(0.06 – 0.13)*
Routine, highly practiced, rapid task involving relatively low level of skill.	**0.02** *(0.007 – 0.045)*
Restore or shift system to original or new state following procedures, with some checking.	**0.003** *(0.0008 – 0.007)*
Completely familiar, well designed, highly practised routine task, oft-repeated and performed by well-motivated, highly trained individual with time to correct failures but without significant job aids.	**0.0004** *(0.00008 – 0.009)*
Respond correctly to system even when there is an augmented or automated supervisory system providing accurate interpretation of system state.	**0.00002** *(0.000006 – 0.00009)*
Miscellaneous task for which no description can be found.	**0.03** *(0.008 – 0.11)*

HEART Error Producing Conditions

Error Producing Condition	Multiplication Factor
Short time available for correction	17
Ambiguity in required standards	5
Poor / ambiguous feedback	4
Little or no independent checking	3
Unclear allocation of responsibility	1.6
Low intrinsic meaning in a task	1.4
High level emotional stress	1.3
Excess team members (per head)	1.03

HEART Generic Violation Behaviors

Generic Violation Behaviours	Nominal error probabilities for females (x 1.4 for males)
Distinctly inconvenient to comply. Potential violator does not feel bound by any implied requirement to comply. Easy to violate. Little likelihood of detection.	0.42
Compliance relatively unimportant. Easy to violate. Little or no inducements to comply.	0.35
Compliance may be fairly important, but chances of detecting violation low. Personal benefits of violating are high and direct.	0.38
Personal benefit in violating, though likelihood of detection is moderate to high. Or else compliance fairly important, but chances of detection low.	0.18
Compliance important, usually legally required, but chances of detection low to moderate.	0.03
No immediate incentive to violate, but likelihood of violation detection moderate to high	0.007
Socially unacceptable, likelihood of detection low and likelihood of unfavourable outcome for violator low.	0.007
Socially unacceptable, chances of detection high and chances of bad outcome high.	0.0001

Data size range	CPU sec	General Comments
Kilobyte ‘000 (10^3)		Who will ever need more than 640KB? (circa 1981)

Data size range	CPU sec	General Comments
Kilobyte ‘000 (10^3)		Who will ever need more than 640KB? (circa 1981)
Megabyte ‘000,000 (10^6)		Stochastic modelling range financial and risk Human Genome 725 Mbytes – 2GB equivalent (as compressed form) Cutting edge Autonomous vehicles about 750 Mb/second

Data size range	CPU sec	General Comments
Kilobyte ‘000 (10^3)		Who will ever need more than 640KB? (circa 1981)
Megabyte ‘000,000 (10^6)		Stochastic modelling range financial and risk Human Genome 725 Mbytes – 2GB equivalent (as compressed form) Cutting edge Autonomous vehicles about 750 Mb/second
Gigabyte ‘000,000,000 (10^9)	PC’s	Movies 2-20 GB.

Data size range	CPU sec	General Comments
Kilobyte ‘000 (10^3)		Who will ever need more than 640KB? (circa 1981)
Megabyte ‘000,000 (10^6)		Stochastic modelling range financial and risk Human Genome 725 Mbytes – 2GB equivalent (as compressed form) Cutting edge Autonomous vehicles about 750 Mb/second
Gigabyte ‘000,000,000 (10^9)	PC’s	Movies 2-20 GB.
Terabyte ‘000,000,000,000 (10^{12})	Tesla AI/AV	Average home HDD’s Tesla AV CPU 36 Trillion ops/second

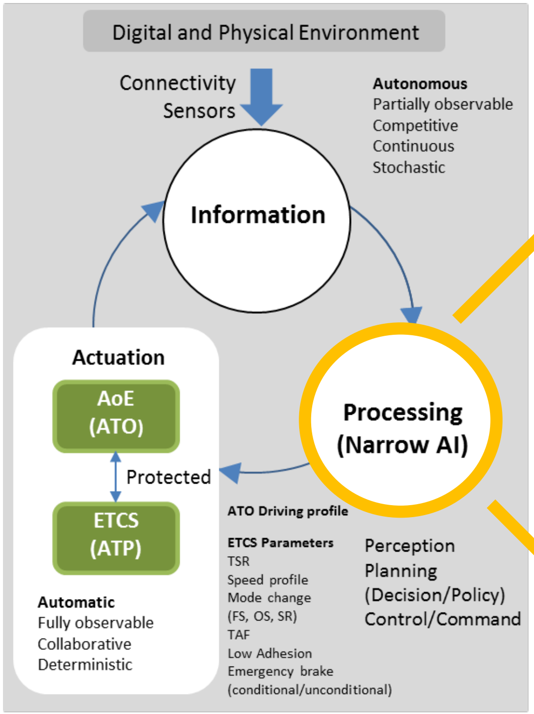
Data size range	CPU sec	General Comments
Kilobyte ‘000 (10^3)		Who will ever need more than 640KB? (circa 1981)
Megabyte ‘000,000 (10^6)		Stochastic modelling range financial and risk Human Genome 725 Mbytes – 2GB equivalent (as compressed form) Cutting edge Autonomous vehicles about 750 Mb/second
Gigabyte ‘000,000,000 (10^9)	PC’s	Movies 2-20 GB.
Terabyte ‘000,000,000,000 (10^{12})	Tesla AI/AV	Average home HDD’s Tesla AV CPU 36 Trillion ops/second
Petabyte (10^{15})	LHC	Facebook generates about 4 Petabytes per day Large Hadron collider 1 petabyte per second when working Estimated human brain capacity 2.5 Petabytes

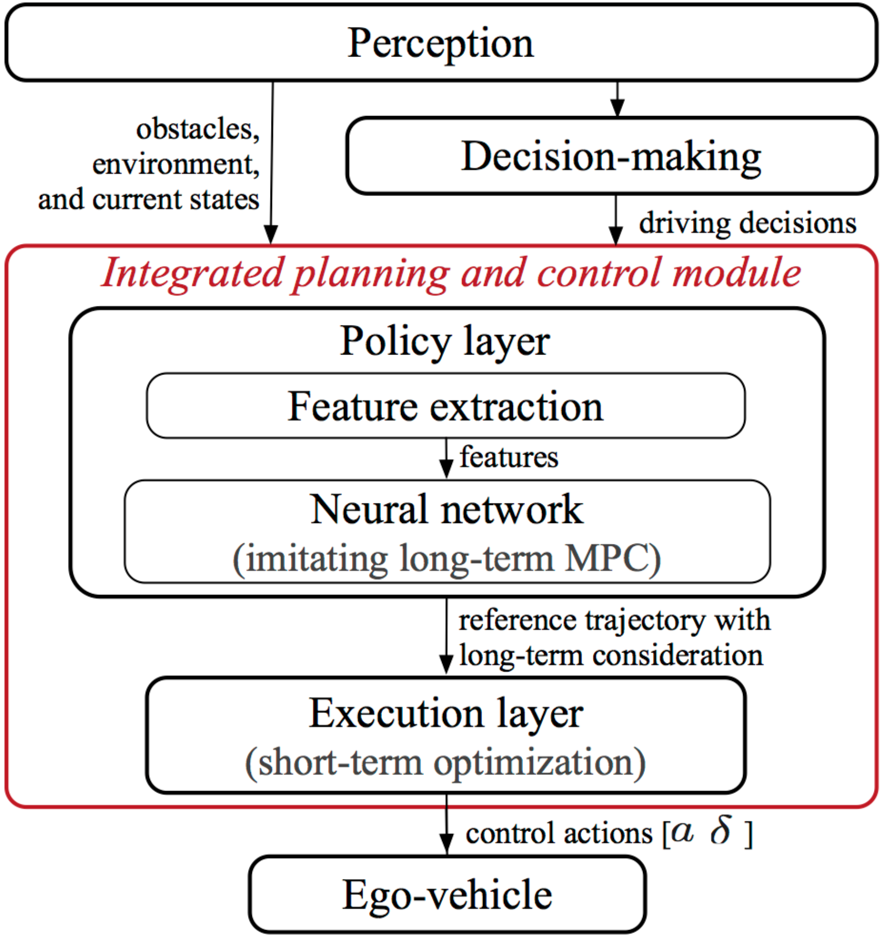
Data size range	CPU sec	General Comments
Kilobyte '000 (10^3)		Who will ever need more than 640KB? (circa 1981)
Megabyte '000,000 (10^6)		Stochastic modelling range financial and risk Human Genome 725 Mbytes – 2GB equivalent (as compressed form) Cutting edge Autonomous vehicles about 750 Mb/second
Gigabyte '000,000,000 (10^9)	PC's	Movies 2-20 GB.
Terabyte '000,000,000,000 (10^{12})	Tesla AI/AV	Average home HDD's Tesla AV CPU 36 Trillion ops/second
Petabyte (10^{15})	LHC	Facebook generates about 4 Petabytes per day Large Hadron collider 1 petabyte per second when working Estimated human brain capacity 2.5 Petabytes
Exabyte (10^{18})	Global	Everyday, globally we create 2.5 exabytes (quintillion) of **data** **6.4 exabits/second for each human brain nerve impulses estimated (2011 – similar ops/second of all computers), genomics generates about 50-100 exabytes per year**

Data size range	CPU sec	General Comments
Kilobyte '000 (10^3)		Who will ever need more than 640KB? (circa 1981)
Megabyte '000,000 (10^6)		Stochastic modelling range financial and risk Human Genome 725 Mbytes – 2GB equivalent (as compressed form) Cutting edge Autonomous vehicles about 750 Mb/second
Gigabyte '000,000,000 (10^9)	PC's	Movies 2-20 GB.
Terabyte '000,000,000,000 (10^{12})	Tesla AI/AV	Average home HDD's Tesla AV CPU 36 Trillion ops/second
Petabyte (10^{15})	LHC	Facebook generates about 4 Petabytes per day Large Hadron collider 1 petabyte per second when working Estimated human brain capacity 2.5 Petabytes
Exabyte (10^{18})	Global	Everyday, globally we create 2.5 exabytes (quintillion) of **data** **6.4 exabits/second for each human brain nerve impulses estimated (2011 – similar ops/second of all computers), genomics generates about 50-100 exabytes per year**
Zettabyte (10^{21})	Global	90% of the data in the world today has been created in the last two years alone. About 2.7 Zettabytes in aggregate (quantity versus quality here) 150 Zettabytes stored in the average human's cells (<u>1.5 Gbytes x 100 trillion cells)</u>

Data size range	CPU sec	General Comments
Kilobyte '000 (10^3)		Who will ever need more than 640KB? (circa 1981)
Megabyte '000,000 (10^6)		Stochastic modelling range financial and risk Human Genome 725 Mbytes – 2GB equivalent (as compressed form) Cutting edge Autonomous vehicles about 750 Mb/second
Gigabyte '000,000,000 (10^9)	PC's	Movies 2-20 GB.
Terabyte '000,000,000,000 (10^{12})	Tesla AI/AV	Average home HDD's Tesla AV CPU 36 Trillion ops/second
Petabyte (10^{15})	LHC	Facebook generates about 4 Petabytes per day Large Hadron collider 1 petabyte per second when working Estimated human brain capacity 2.5 Petabytes
Exabyte (10^{18})	Global	Everyday, globally we create 2.5 exabytes (quintillion) of **data** **6.4 exabits/second for each human brain nerve impulses estimated (2011 – similar ops/second of all computers), genomics generates about 50-100 exabytes per year**
Zettabyte (10^{21})	Global	90% of the data in the world today has been created in the last two years alone. About 2.7 Zettabytes in aggregate (quantity versus quality here) 150 Zettabytes stored in the average human's cells (1.5 Gbytes x 100 trillion cells)
Yottabyte (10^{24})	Future Global	Logically in this space within about 3-4 years

Typical AI Vehicle System





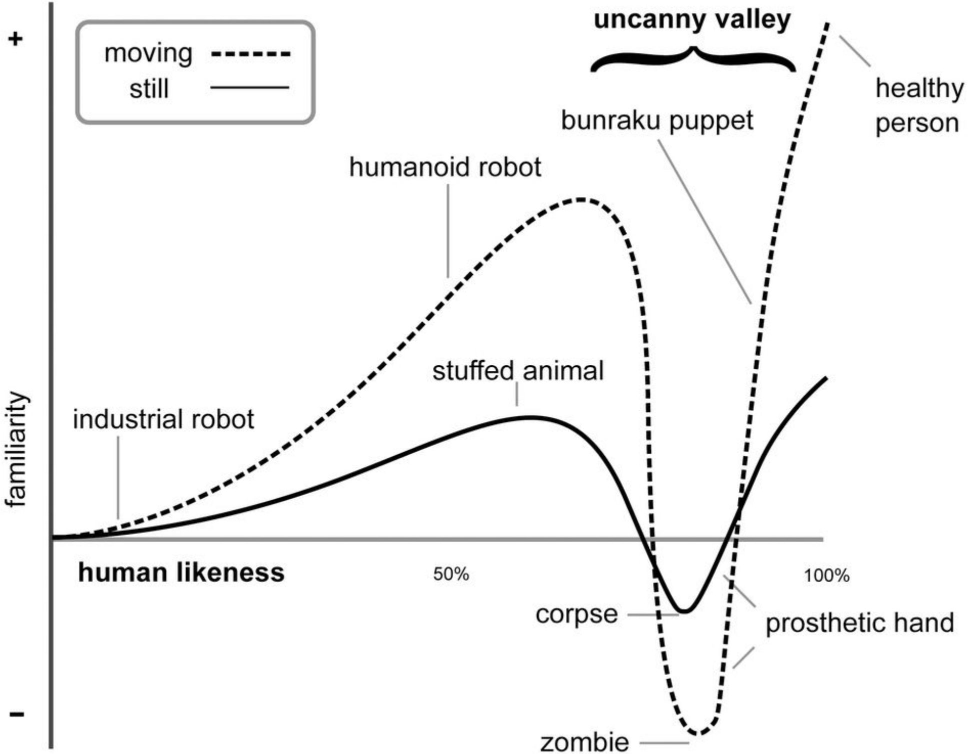
Organisational impacts of AI

- Is your organisation a Gigabyte or a Terrabyte organisation?
- Where is your thinking at?
- Transition to a Petabyte and Exabyte mode of thinking (Facebook, Microsoft, Google and similar territory)
- Stakeholder Messaging
- Education and upskilling
- Organisations need to develop Stakeholder KPI’s
    - Journey
    - Content for events for context
    - Pre inform so less infotainment factor for the media

Societal Expectations Regarding Confidence – reference points

- SIL 4 10,000 to 100,000 (10^4 to 10^5) for rail and nuclear safety
- Proof of new particle discovery Higgs Boson – 5 σ
- Privacy – we can be personally lax, but demand very high sigma externally
- Safety – Highly trained humans driving cars, planes and trains
- Capability – 1 -2 σ better than we do on average
- Honesty/Integrity – Our current human governance systems in organisations and governments with IT and controls (already implicit AI)
- Health Professionals – human diagnosis and treatment with extensive IT and machine support.

Uncanny Valley



Deeper Uncanny Thinking

Coherent Extrapolated Volition (CEV) is intended to be what humanity objectively would want, all things considered, but it can only be defined relative to the psychological and cognitive qualities of present-day, un-extrapolated humanity.

Our feeling of caution or discomfort is to do with not only the visual appearance, but the lack of knowledge of what similarity it will have to our volition.

Therefore, assurance has to be centred around validation and verification of volition of key values and principles (4V’s).

Mountain of Caution

ISO/IEC SC42

Standards for Management of Artificial Intelligence – Early 2018

Aust Stds IT-043 (Mirror)

Standards for Management of Artificial Intelligence – March 2019

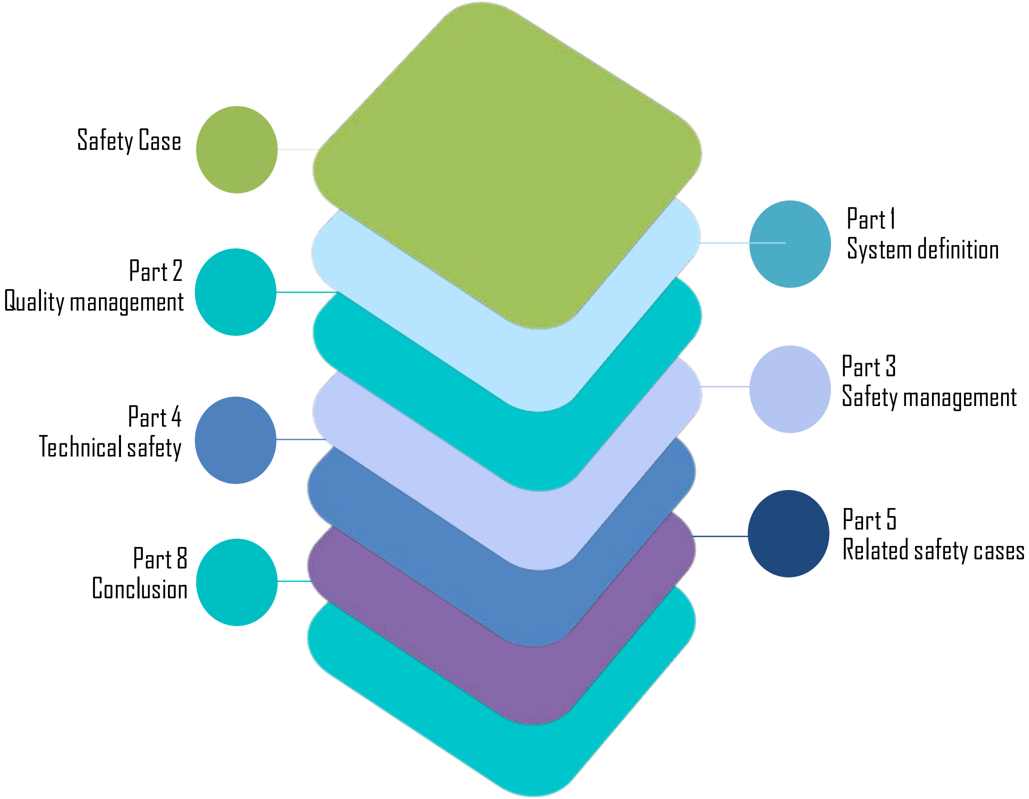
ISO/IEC JTC 1/SC 42 - Artificial intelligence

- Formed in 2018
- International response
- June 2019 – Australian discussion paper
- Sep 2019 – feedback to IT-043 members from Australian survey

ISO/IEC JTC 1/SC 42 - Artificial intelligence

Project	Focus area
ISO/IEC AWI TR 20547-1	Information technology – Big data reference architecture – Part 1: Framework and application process
ISO/IEC DIS 20547-3	Information technology – Big data reference architecture – Part 3: Reference architecture
ISO/IEC WD 22989	Artificial intelligence – Concepts and terminology
ISO/IEC WD 23053	Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)
ISO/IEC AWI 23894	Information Technology – Artificial Intelligence – Risk Management
ISO/IEC NP TR 24027	Information technology – Artificial Intelligence (AI) – Bias in AI systems and AI aided decision making
ISO/IEC PDTR 24028	Information technology – Artificial Intelligence (AI – Overview of trustworthiness in Artificial Intelligence
ISO/IEC NP TR 24029-1	Artificial Intelligence (AI) – Assessment of the robustness of neural networks – Part 1: Overview
ISO/IEC NP TR 24030	Information technology – Artificial Intelligence (AI) – Use cases
ISO/IEC NP TR 24368	Information technology – Artificial intelligence – Overview of ethical and societal concerns
ISO/IEC NP TR 24372	Information technology – Artificial intelligence (AI) – Overview of computational approaches for AI systems
ISO/IEC NP 38507	Information technology – Governance of IT – Governance implications of the use of artificial intelligence by organizations

Traditional Safety Case for Rail, Oil and Gas, & other Facilities



Intelligence – data and AI engine

Data – suitable for level of performance required in the application

Sensors required based on needs

Rich field of data

Changing nature of data

Different types of statistical bias (including over 40 types of human bias)

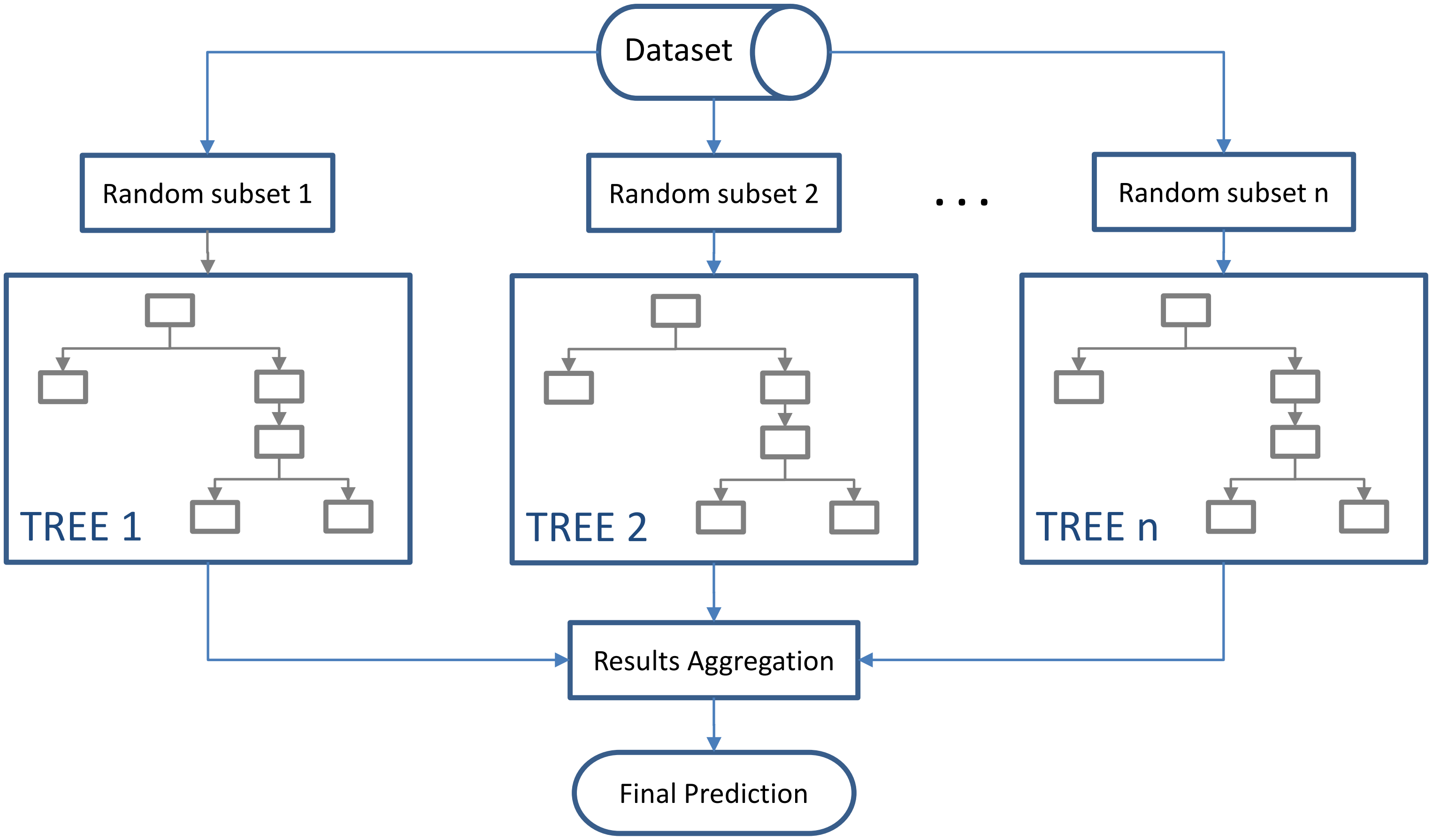
Dropout

Too simple

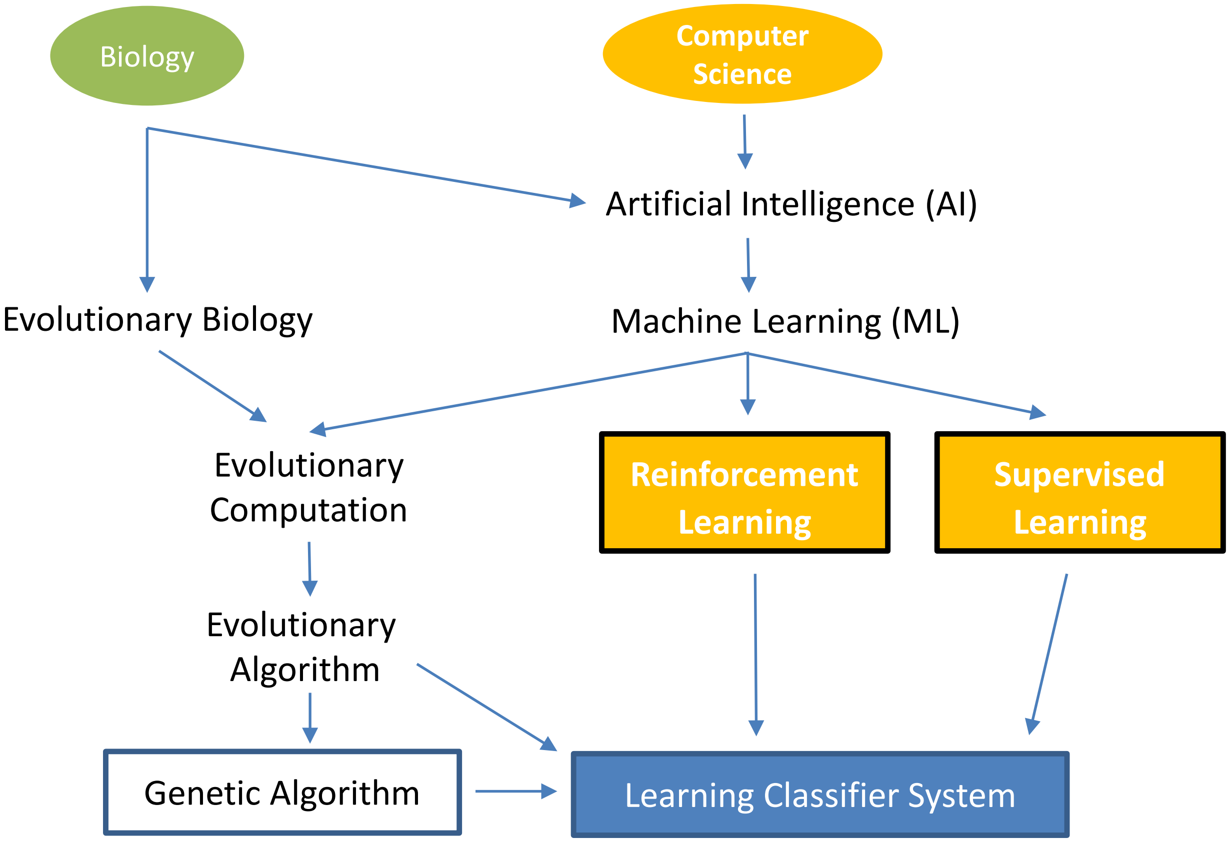
OverFitting, Regularization

Normalization

Ensemble Learning – used with AVs



Ensemble Learning with Genetic algorithm – used with AVs



Adversarial Learning

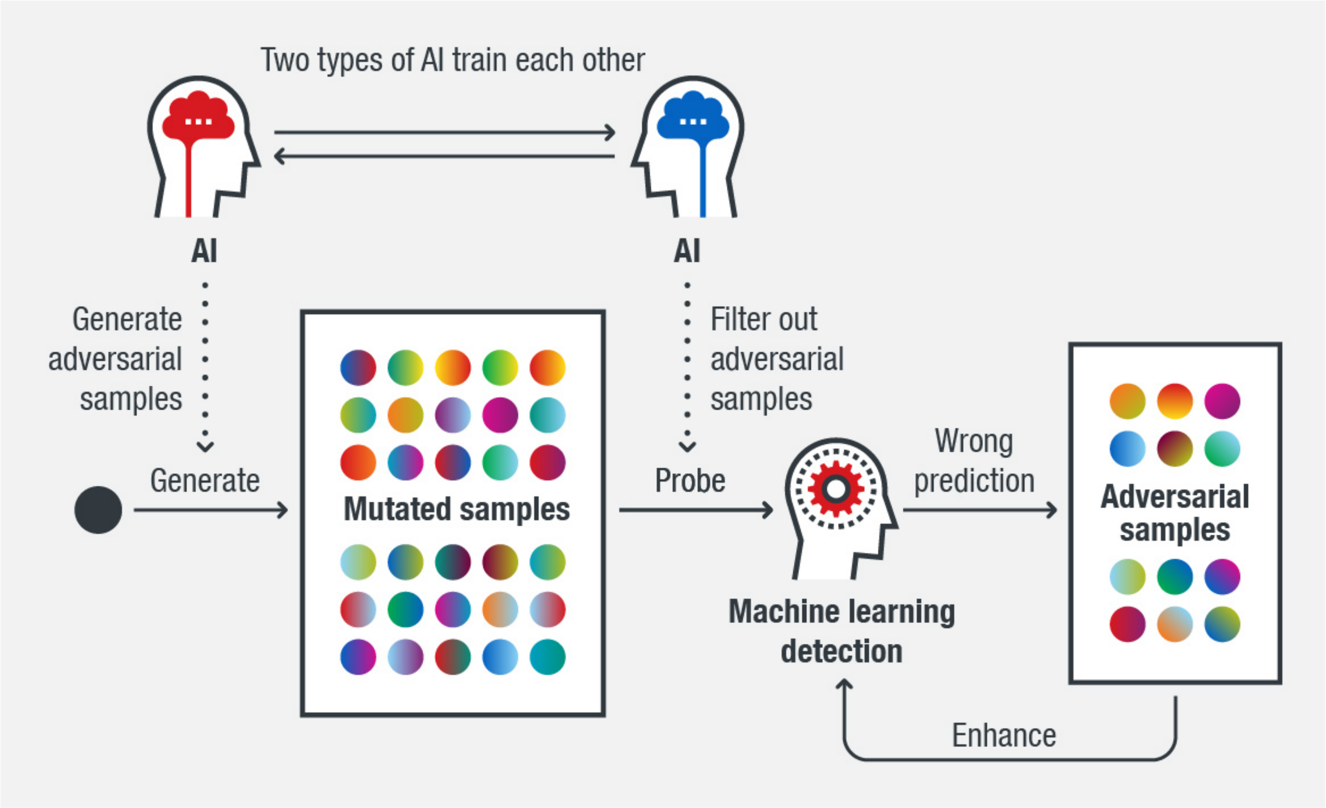
Poisoning Attacks

- An ANN attacker poisons the training data by injecting carefully designed samples to eventually compromise the whole learning process.

Evasion Attacks

- An ANN attacker modifies samples at test time to evade detection; that is, to be misclassified as legitimate.

Adversarial Learning



Conclusions

- AI in train automation is happening.
- Potential for safety, operational and financial benefit is high.
- There are also many risks that need to be managed.
- Validation and Verification approaches need updating for non deterministic complex systems.
- Higher standards of safety will likely be expected
- Human Factors, risk management and stakeholder engagement are very important.
- ETCS/CBTC provides foundational benefit for all levels of automation.
- There is potential of new Human Factors risks in a region of higher levels of automation as evidenced from aviation, in vehicles, and similarly for trains.
- Standards need updating, i.e. functional safety, new generation AI standards, IEEE standards,

GHD

DIGITAL TRAIN CONTROL
(ETCS, CBTC, PTC)

GHD
Advisory

Autonomous system model