Digital Privacy

Włodzisław Duch

Neurocognitive Laboratory, Center for Modern Interdisciplinary Technologies,
& Dept. of Informatics, Faculty of Physics, Astronomy & Informatics,
Nicolaus Copernicus University, Toruń, Poland

Google: Wlodzislaw Duch

IEEE SMC, Prague 10/2022

EU main documents

Coordinated Plan on Artificial Intelligence 12/2018

2/2020 White Paper: On Artificial Intelligence - A European approach to excellence and trust.
Commission supports a regulatory and investment oriented approach, promoting the uptake of AI and addressing the risks associated with certain uses of this new technology. AI ecosystem based on trustworthiness, European values and rules, **'ecosystem of excellence'.**

H2020 Call on European Network of Artificial Intelligence Excellence Centres: Information and Brokerage day, Brussels 5/2019, included presentations of many ideas for AI labs.

Ethical issues dominate … European approach to AI centers on excellence and trust, includes legal framework, AI liability directive (9/22), and proposal for product liability directive. But technological issues are first:

Pestian J.P, Itert L, Andersen C, Duch W, *Preparing Clinical Text for Use in Biomedical Research.* Journal of Database Management 17(2), 1-11, 2006.

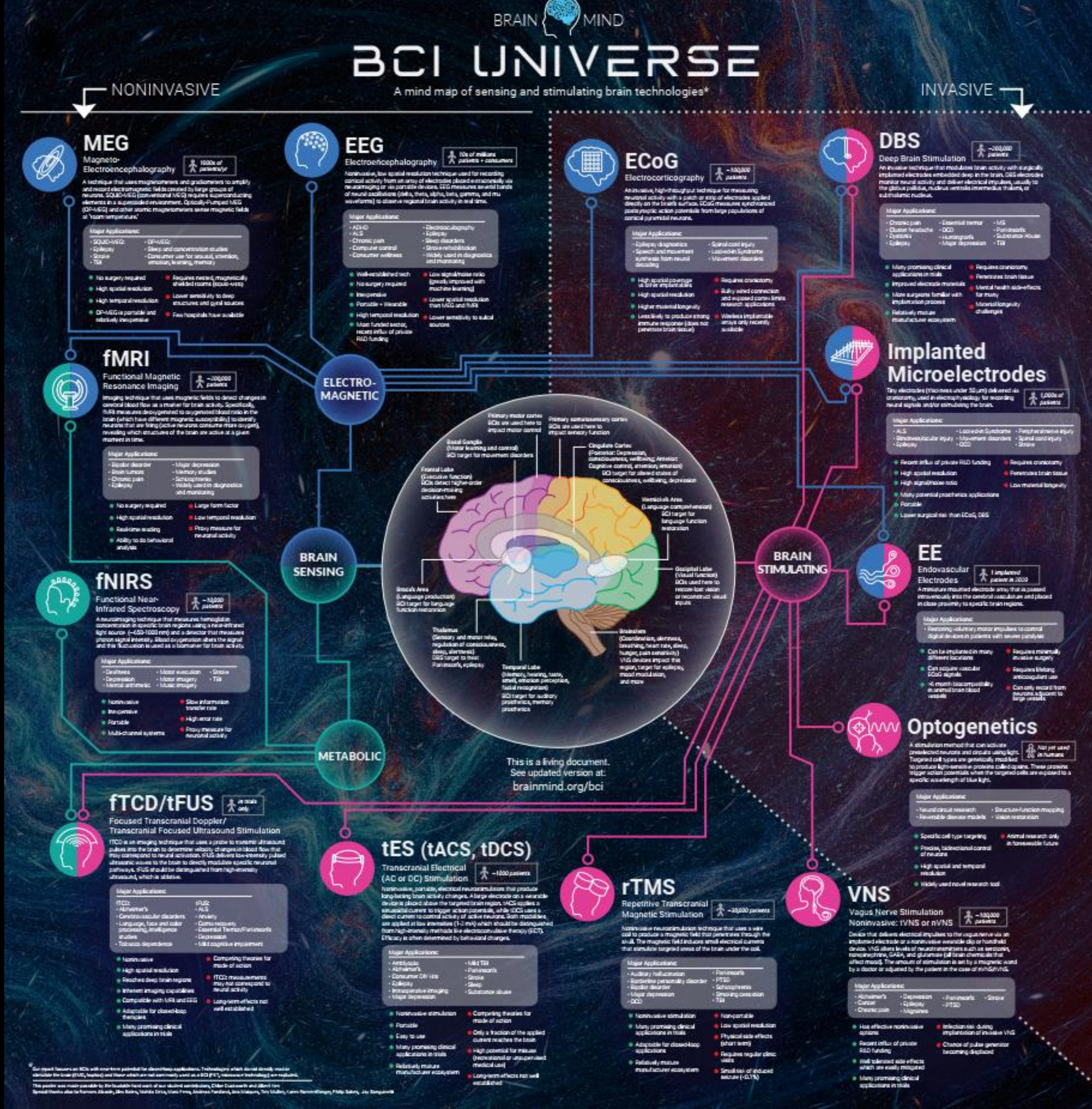
EU AI Initiatives 2018-2020

High-Level EU Expert Group on Artificial Intelligence,
Coordinated Plan on Artificial Intelligence.

The European AI-on-Demand Platform and Ecosystem,
AI Digital Innovation Hubs (AI DIHs), Digital Innovation Hub Healthcare Robotics

Building Trust in Human-Centric AI - Horizon 2020, invested 9/2020 €50 million in ELISE, Clair, AI4Media, HumanE-AI-Net. Ethic guidelines conform to the **Assessment List for Trustworthy Artificial Intelligence (ALTAI)**

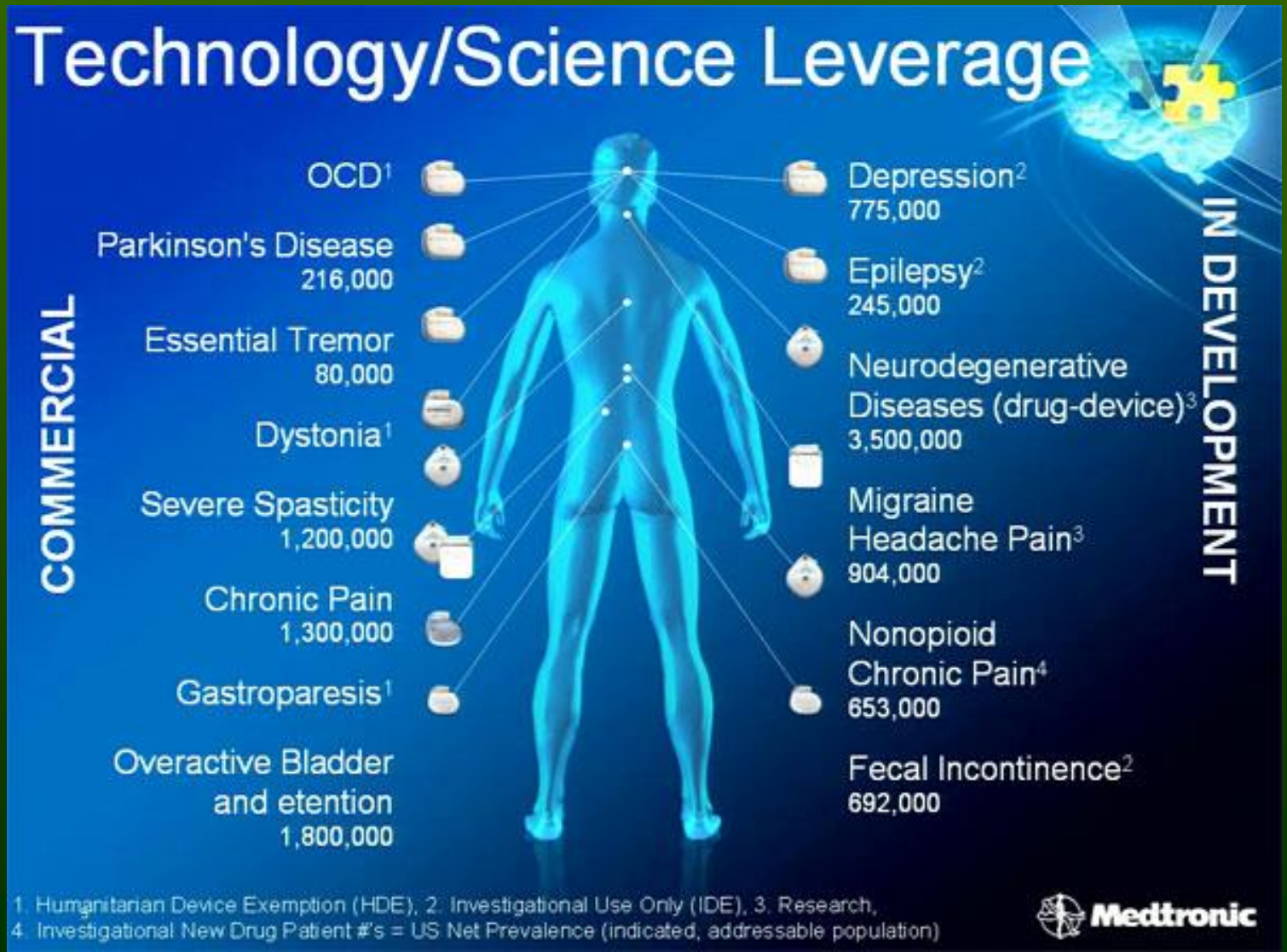
1. Human agency and oversight,
2. Technical robustness and safety, secure
3. Privacy and data governance,
4. Transparency, explainable,
5. Diversity, non-discrimination and fairness,
6. Societal and environmental wellbeing,
7. Accountability.


BRAIN MIND
BCI UNIVERSE
A mind map of sensing and stimulating brain technologies*
NONINVASIVE
INVASIVE
MEG
Magneto-Electroencephalography
EEG
Electroencephalography
ECoG
Electrocorticography
DBS
Deep Brain Stimulation
fMRI
Functional Magnetic Resonance Imaging
ELECTRO-MAGNETIC
Implanted Microelectrodes
BRAIN SENSING
BRAIN STIMULATING
fNIRS
Functional Near-Infrared Spectroscopy
EE
Endovascular Electrodes
METABOLIC
Optogenetics
This is a living document.
See updated version at:
brainmind.org/bci
fTCD/tFUS
Focused Transcranial Doppler/
Transcranial Focused Ultrasound Stimulation
tES (tACS, tDCS)
Transcranial Electrical (AC or DC) Stimulation
rTMS
Repetitive Transcranial Magnetic Stimulation
VNS
Vagus Nerve Stimulation
Noninvasive: tVNS or nVNS

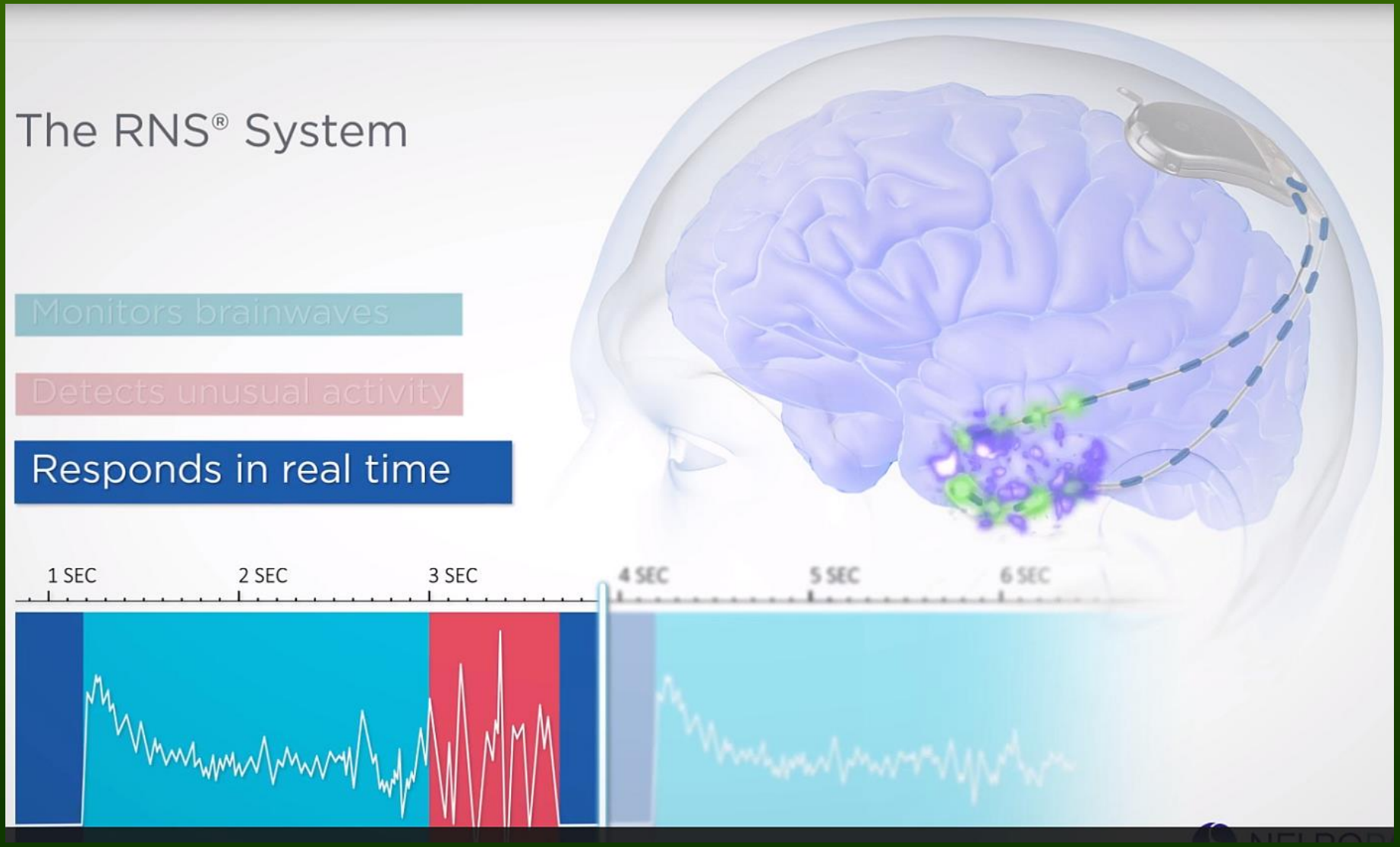
Neuromodulation

Cochlear implants are common, deep implants stimulate brain structures, not only for deficits of perception, but to regulate cortical neural processes.

Market:
10B$ (2021),
25B$ in 2027.



Epilepsy



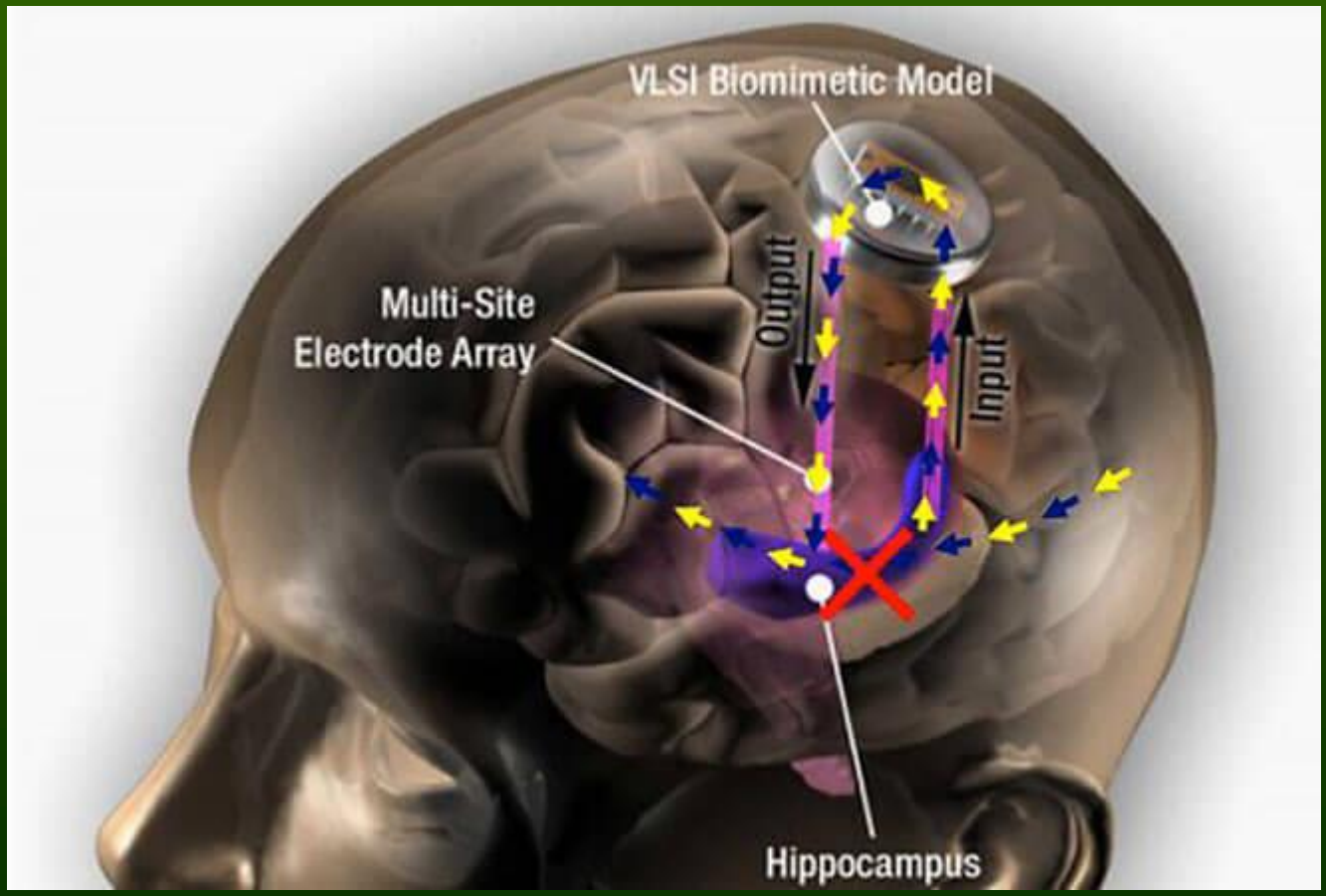
The neurostimulator and detector stops attacks of drug-resistant epilepsy before cramps occur. About 1% of people in the world have epilepsy.

Memory implants – Inception?

Tests on rats, monkeys, and in 2017 on 20 humans gave an improvement in memory by 30% (on rats by 35%). Ted Berger (USC, Kernel) : There are good reasons to believe that the integration of memory with electronics is possible.

DARPA: Restoring Active Memory (RAM) program, for people with brain damage (TBI), should be non-invasive.

Neurofeedback + closed-loop neurostimulation.



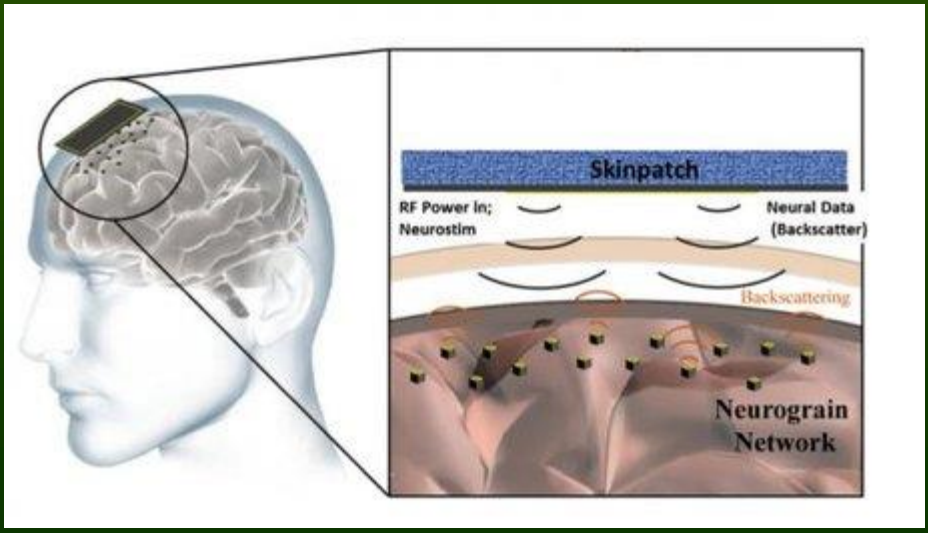
A million nanowires in the brain?

DARPA initiative: **Neural Engineering System Design (NESD)** and other projects.

An interface that reads the impulses of 10^6 neurons, stimulates 10^5 neurons, simultaneously reads and stimulates 10^3 neurons.

DARPA awarded grants to research groups for projects under the program Electrical Prescriptions (ElectRx), whose aim is to develop BCBI systems modulating the activity of peripheral nerves for therapeutic purposes.

Neural dust – microscopic wireless sensors in the brain.
Elon Musk and the much-heralded technology neuralink (neural lace).



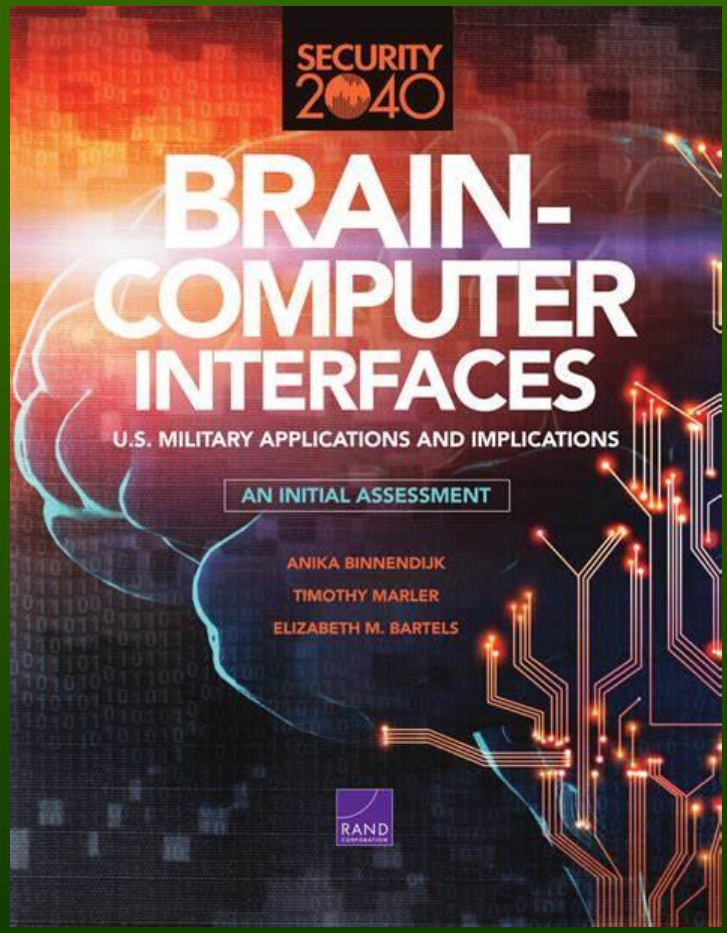
Privacy Threats

BCI – passive, reading brain states, or active, neuromodulation, deep brain stimulation, epilepsy, prosthetics, cochlear implants ...

Risks: intercepting BCI signals, jamming, accessing operator's emotional/cognitive states, controlling neuromodulation ...

Implement National Academy of Sciences and European Union ethical recommendations in development and implementation:

1. consent of patients/service members,
2. health implications for invasive BCI,
3. enhanced human performance considerations,
4. **potential risks to privacy**.



A Binnendijk, T Marler, EM. Bartels
U.S. Military Applications and Implications, An Initial Assessment, RAND 2020

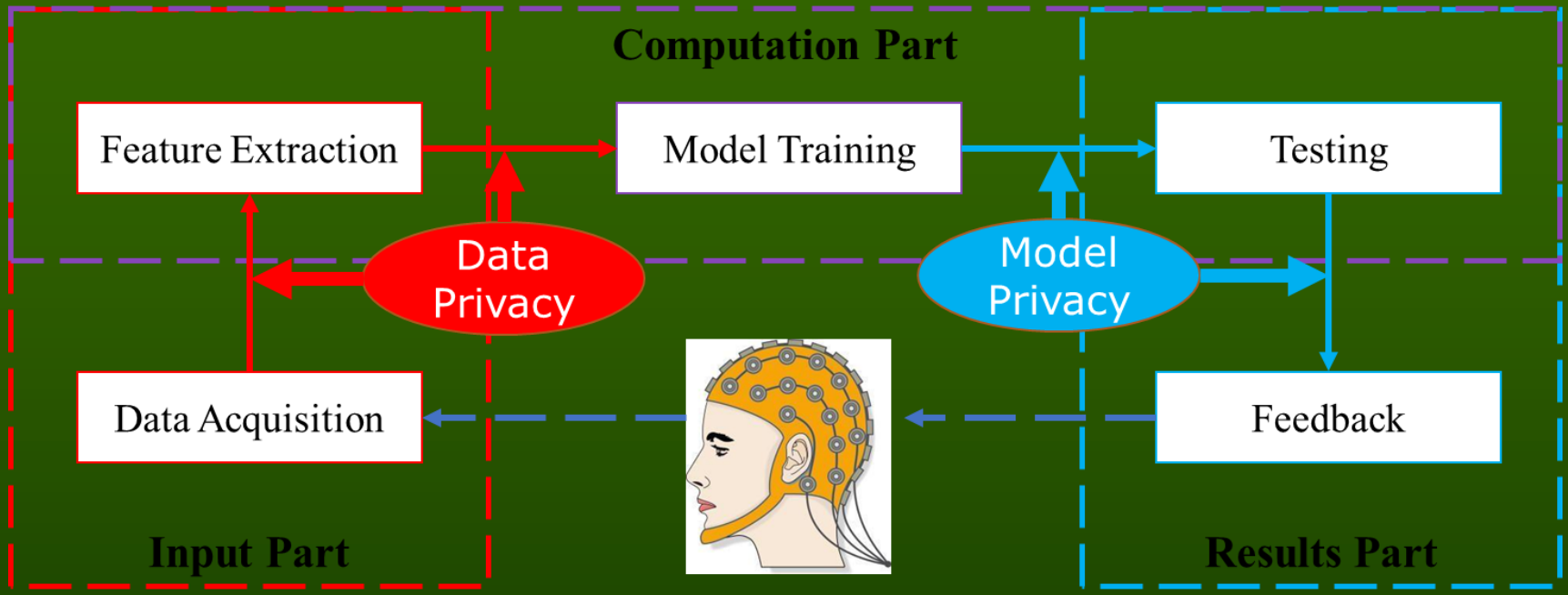
BCI Toolbox for National Security Game

1) Human-machine decisionmaking - Immediate transfer of operational risk • Faster decisions to deploy weapons • Shorter preparation cycle with faster feedback, from occurrences in battlespace • Increased speed and accuracy of targeting • Augmented AI systems
2) Human-machine direct system control - Transfer basic commands to systems • Increase situational awareness and reaction
3) Human-to-human communication/ management - Transfer basic commands between individuals
4) Monitor performance - individual & group cognitive workload, stress
5) Enhance cognitive performance - Regulate stress • Increase focus and alertness
6) Enhance physical performance - Improved strength augmentation • Improved sensory capabilities
7) Training - Increased learning retention • Deployable training devices • Adaptive individualized training • Immediate effective assessment

Private Information in BCIs

K. Xia, W. Duch, Y. Sun, K. Xu, W. Fang, H. Luo, Y. Zhang, D. Sang, D. Wu, X. Xu, F-Y Wang, Privacy-Preserving Brain-Computer Interfaces: A Systematic Review, IEEE Trans. on Computational Social Systems, 2022 (slides from Dongrui Wu)

Privacy Threats in BCIs



- **Input part**: Data acquisition and feature extraction
- **Computation part**: Model training, additional feature extraction
- **Results part**: Uses its own data for testing.
- A closed-loop BCI system maps results part back to input part, user info.
- **Privacy threats:** data transmission from the input part to the computation part, or during model delivery in the test phase.

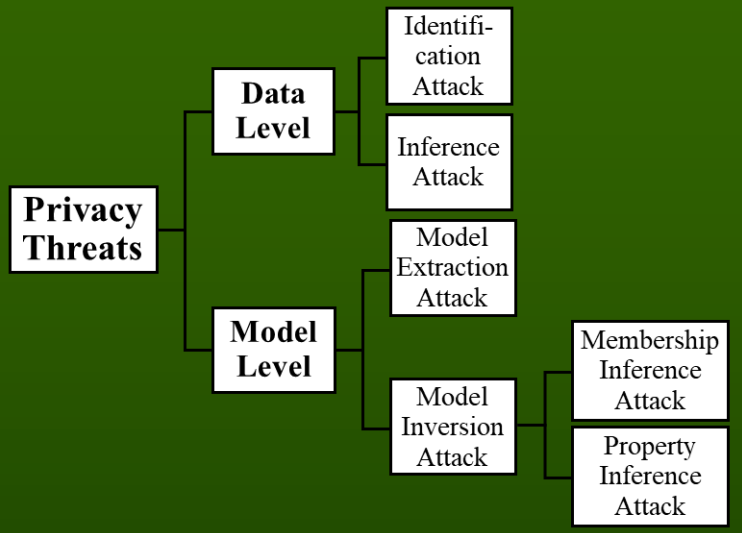
Privacy Threat Types in BCIs

Data-level privacy threats use the information in published data, e.g., raw data, identity, statistical properties, etc.

- ✓ **Identification Attacks**: Infer the identity of the source from the data obtained.
- ✓ **Inference Attacks:** Perform reasoning or learning on data to obtain hidden sensitive information

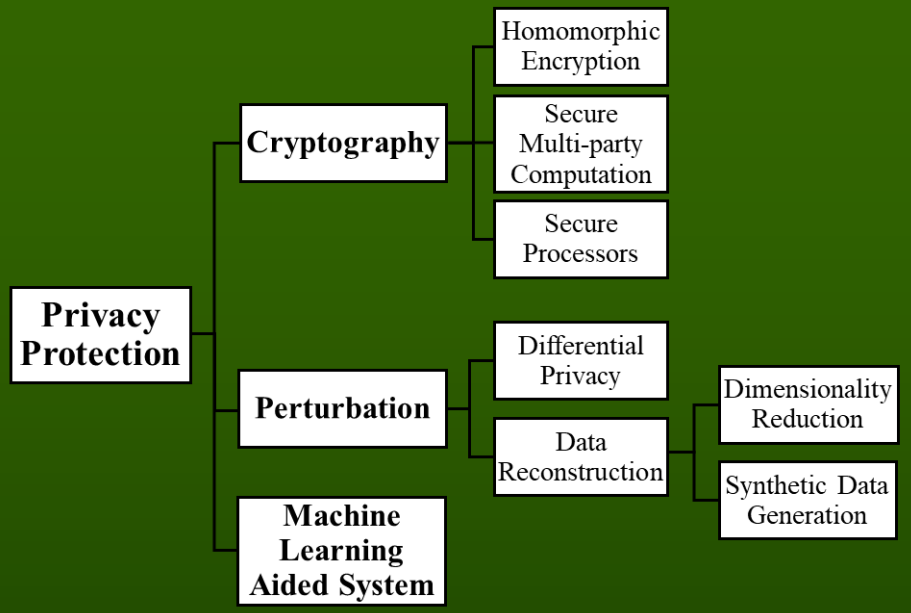
Model-level privacy threats try to extract private information, including its structure, parameters, training algorithms, training data, etc.

- ✓ **Model Extraction Attack:** continuously query the model to obtain its outputs, use model responses to replicate model structure, parameters, training data, etc. or to imitate model functionalities.
- ✓ **Model Inversion Attack:** Infer some sensitive information of the training data from the model.



Privacy Protection Approaches in BCIs

- **Cryptography**: homomorphic encryption, secure multi-party computation, secure processors edge computing.
- **Perturbation:** Resists attacks by adding certain noise or transforming the original data while maintaining the data utility for downstream tasks.
- **ML + explanation aided systems**: help people better understand privacy policies, and inform them about the privacy risks when making privacy decisions.

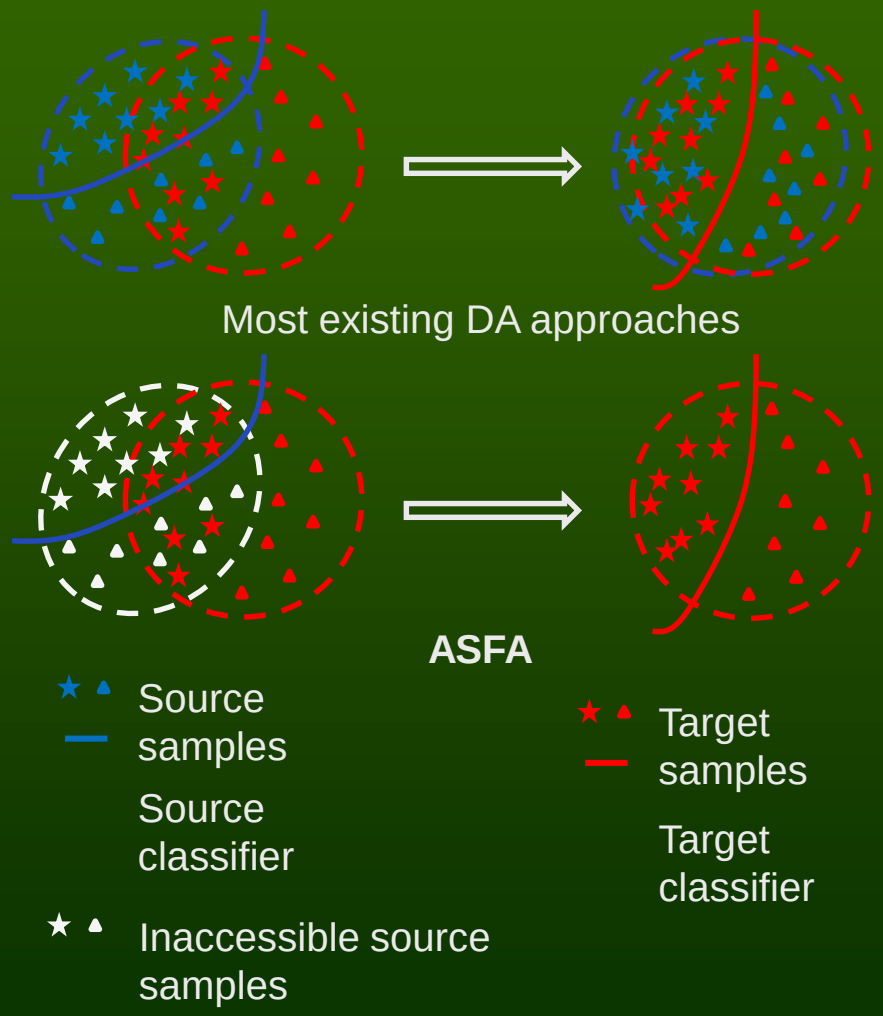


Augmentation-based Source-Free Adaptation (ASFA)

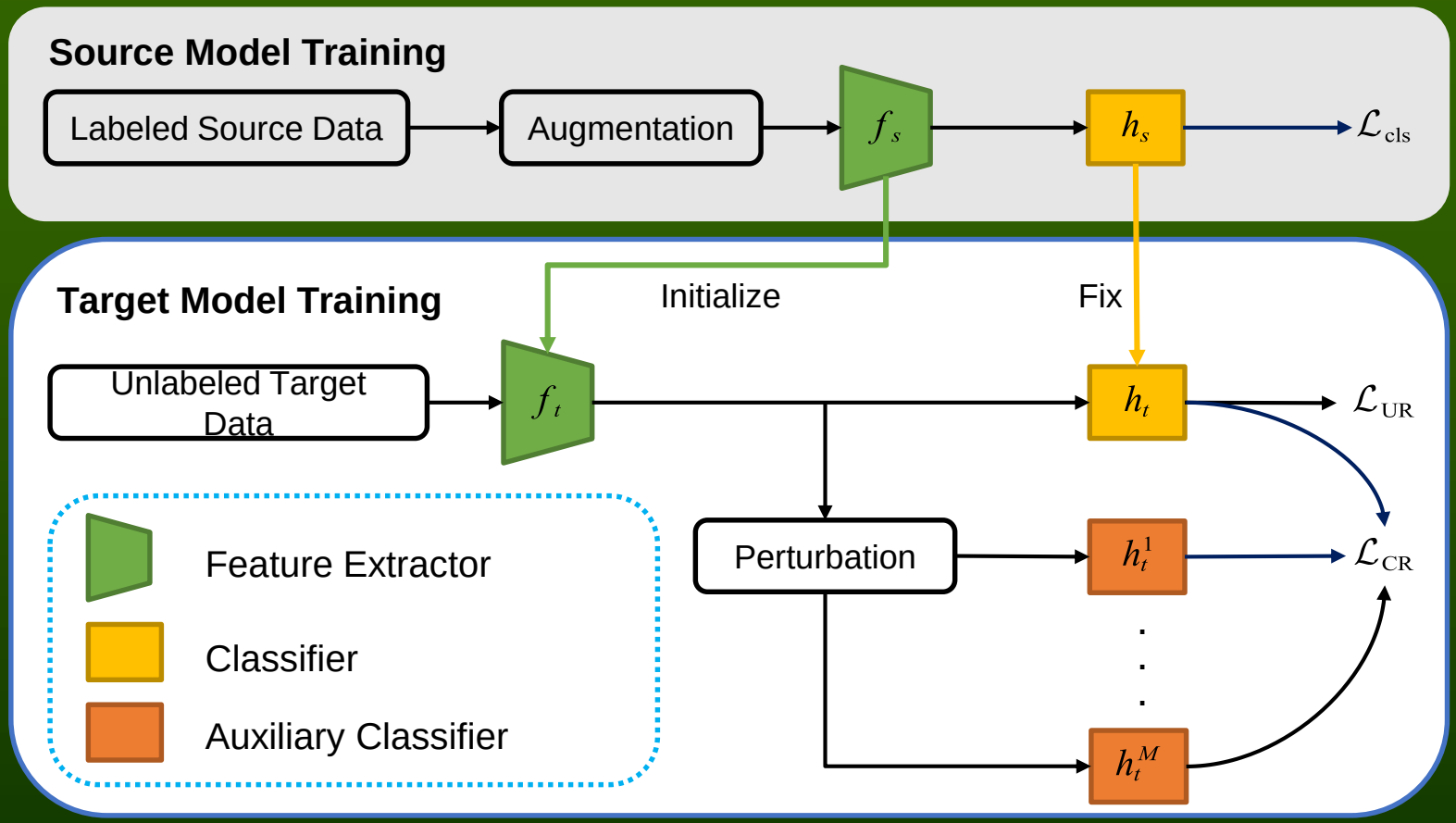
- ASFA consists of two parts:

Source model training, where a novel data augmentation approach for MI EEG signals is proposed to increase the source model generalization ability, by making use of the redundancy of EEG channels.

Target model training, where the target model is initialized from the source model, and trained through uncertainty reduction for reducing the domain shift and consistency regularization for model robustness.



Overview of ASFA



K. Xia, W. Duch, Y. Sun, K. Xu, W. Fang, H. Luo, Y. Zhang, D. Sang, D. Wu, X. Xu, F-Y Wang, Privacy-Preserving Brain-Computer Interfaces: A Systematic Review, IEEE Trans. on Computational Social Systems, 2022 (slides from Dongrui Wu).