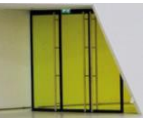
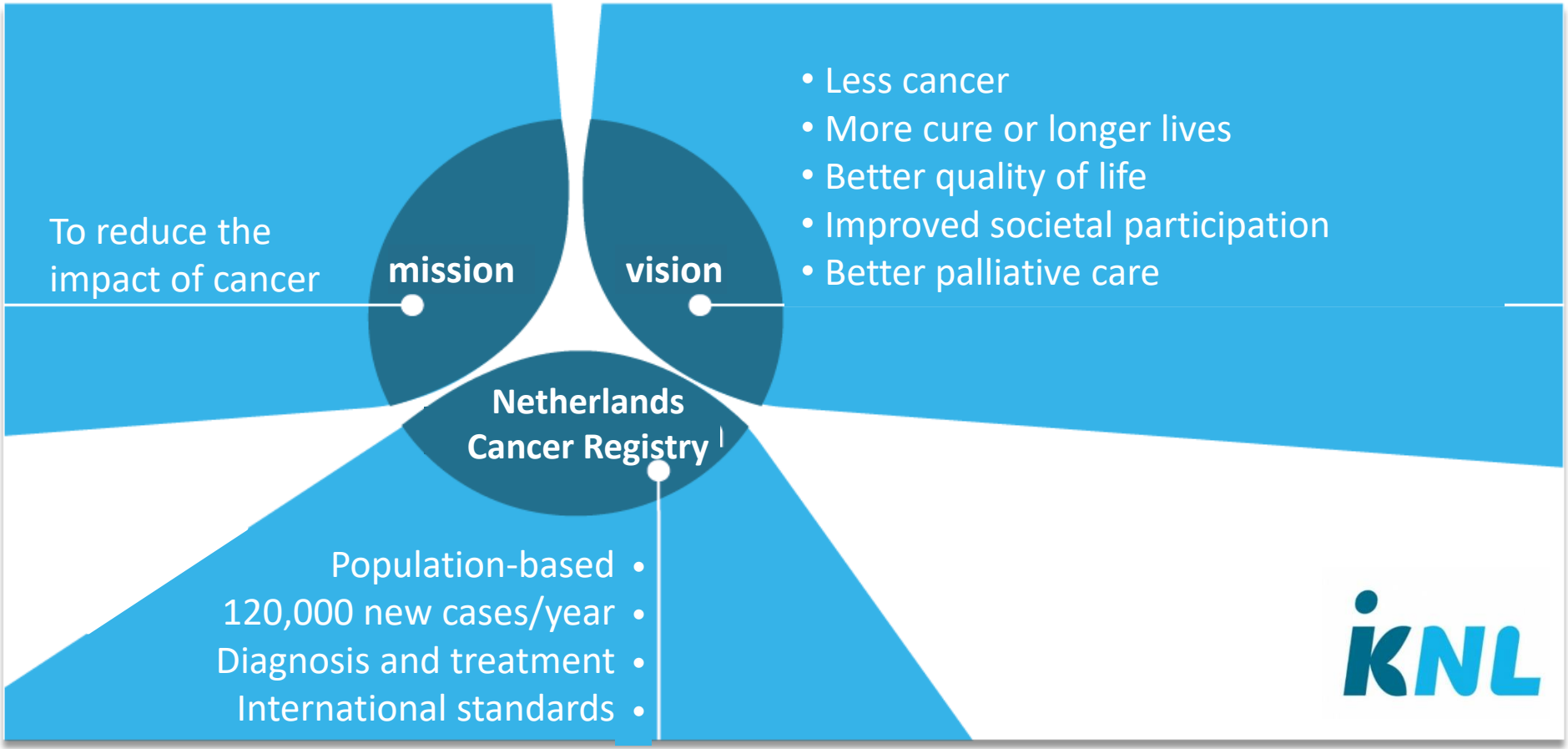
# Enabling privacy-preserving analyses on federated healthcare data.

Gijs Geleijnse (IKNL) , Daniël Worm (TNO)

integraal kankercentrum Nederland

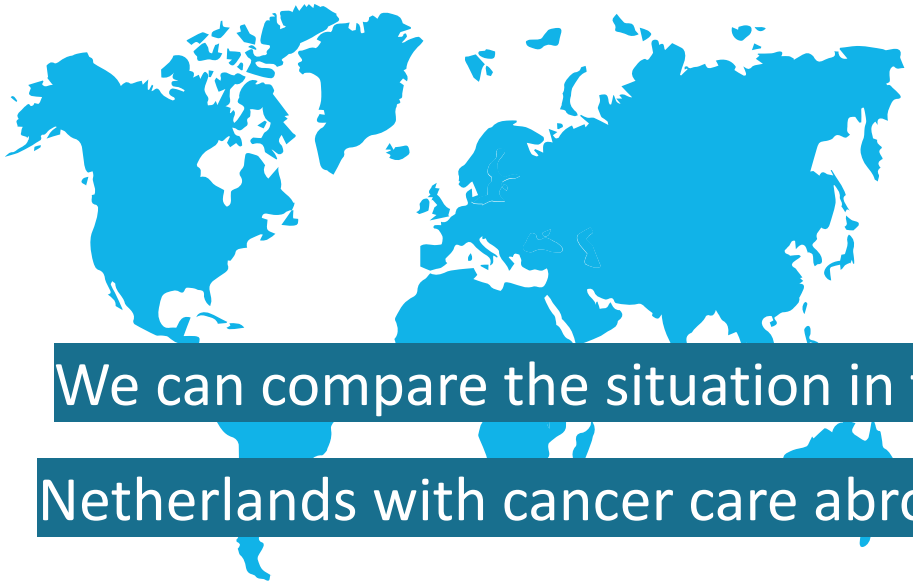
TNO




To reduce the impact of cancer
mission
vision
• Less cancer
• More cure or longer lives
• Better quality of life
• Improved societal participation
• Better palliative care
Netherlands Cancer Registry
Population-based •
120,000 new cases/year •
Diagnosis and treatment •
International standards •
iKNL

# Combining data helps!

Other data sets can give extra information about patients (in the Netherlands)

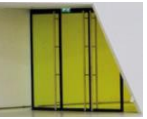


We can compare the situation in the Netherlands with cancer care abroad

# Traditional solution: bringing data together

Researcher processes data on central database

All parties provide their data to the central database.



# Traditional solution: disadvantages

Security

Laws & regulation

Privacy

Loss of control

# Proposed Solutions – Federated Learning and Multi-Party Computation


Bring algorithms to the data

Maximize the potential of multiple datasets without sharing sensitive data

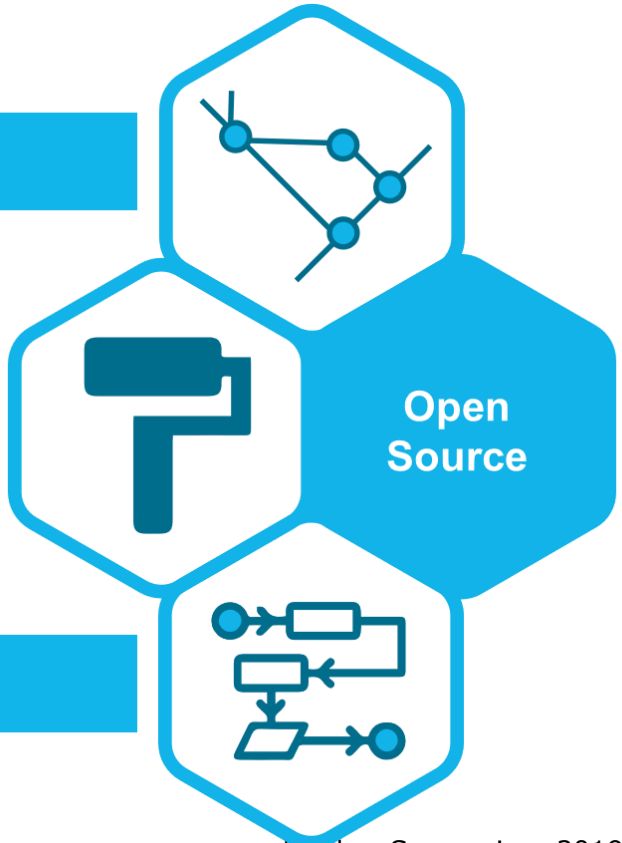
Minimize data leaks and privacy risks

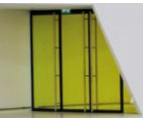
# Separation of concerns

Infrastructure

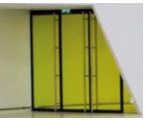
Data format

Algorithms



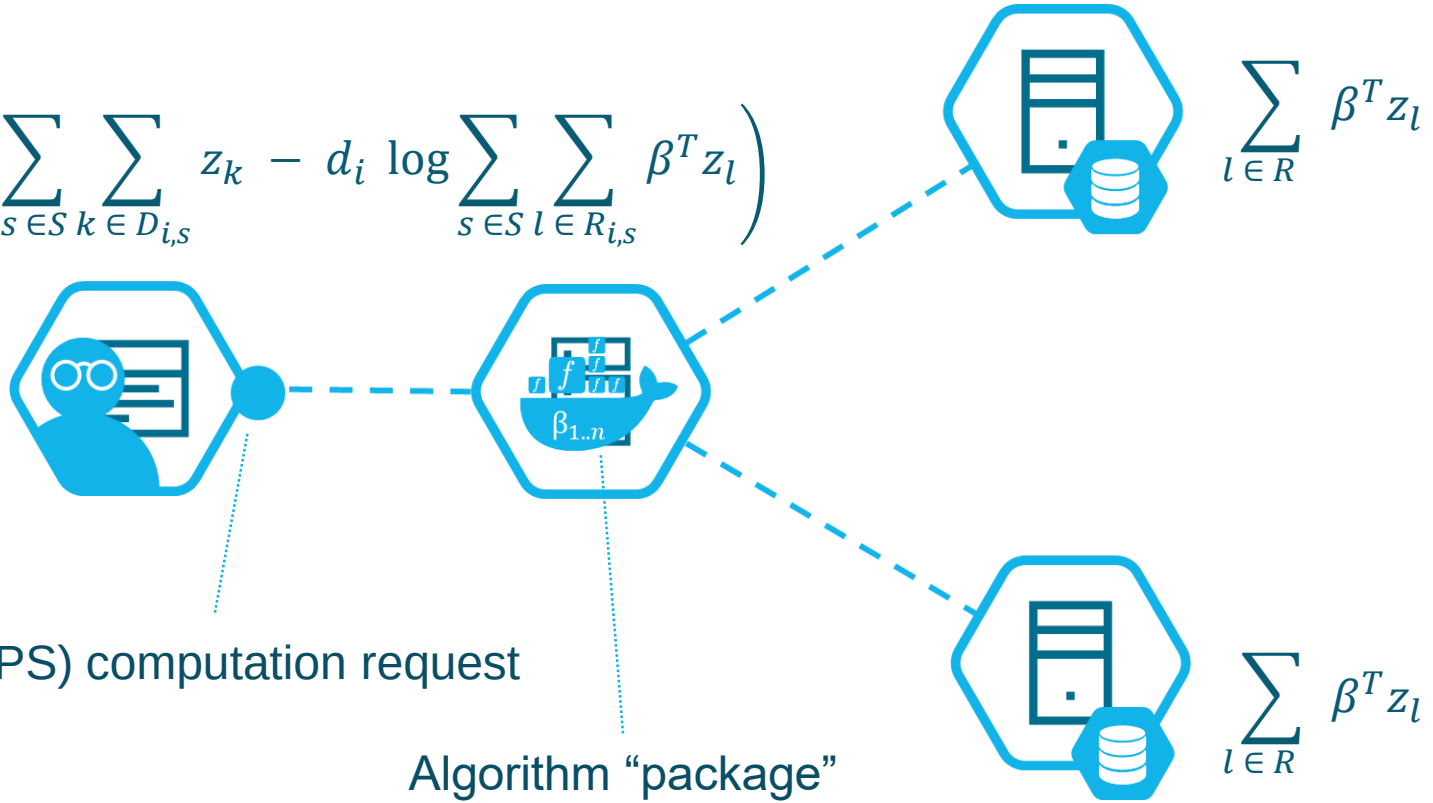



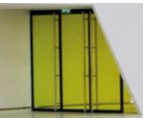
# Horizontally partitioned data



# Vantage infrastructure

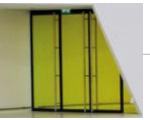
$$l(\beta) = \sum_{i=1}^{D} \beta^T \left( \sum_{s \in S} \sum_{k \in D_{i,s}} z_k - d_i \log \sum_{s \in S} \sum_{l \in R_{i,s}} \beta^T z_l \right)$$



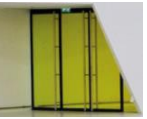



# Enabling international cancer registry research

| | Netherlands | | Taiwan | | Combined | |
|---|---|---|---|---|---|---|
| | **HR** | **95% CI** | **HR** | **95% CI** | **HR** | **95% CI** |
| **Geography** | | | | | | |
| Taiwan | | | | | Reference | |
| The Netherlands | | | | | 1.03 | 0.98 – 1.09 |
| | | | | | | |
| Age | 1.04 | 1.04 – 1.05 | 1.02 | 1.02 – 1.02 | 1.03 | 1.03 – 1.03 |
| **Gender** | | | | | | |
| Female | Reference | | Reference | | Reference | |
| Male | 1.30 | 1.21 – 1.39 | 1.19 | 1.11 – 1.27 | 1.21 | 1.15 – 1.26 |
| **Period of Diagnosis** | | | | | | |
| 2004 - 2007 | Reference | | Reference | | Reference | |
| 2008 - 2011 | 0.93 | 0.86 – 1.01 | 0.81 | 0.78 – 0.85 | 0.84 | 0.81 – 0.88 |
| 2012 - 2016 | 0.86 | 0.78 – 0.93 | 0.66 | 0.63 – 0.69 | 0.71 | 0.69 – 0.74 |
| **Stage** | | | | | | |
| I | Reference | | Reference | | Reference | |
| II | 1.38 | 1.25 – 1.52 | 1.39 | 1.30 – 1.48 | 1.35 | 1.28 – 1.43 |
| III | 1.58 | 1.41 – 1.78 | 2.01 | 1.87 – 2.15 | 1.88 | 1.77 - 2.00 |
| IVA | 2.73 | 2.45 – 3.03 | 3.28 | 3.07 – 3.49 | 3.10 | 2.93 – 3.27 |



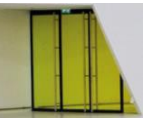
# Vertically Partitioned Data



# CONVINCED

- New collaboration project between IKNL and TNO.

- Goal: enable secure learning of survival analysis models on vertically partitioned data sets.
  - Making use of MPC
  - Jointly developing open source solutions.

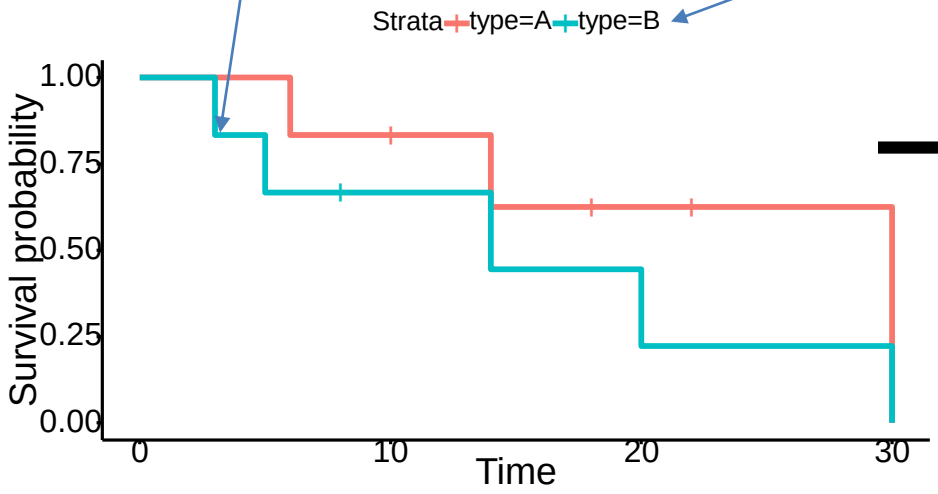


# Kaplan-Meier

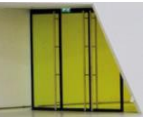
- Non-parametric method to estimate survival curves

Time of death or time of censoring known per patient

Patient type




Statistical difference of survival rates between the types

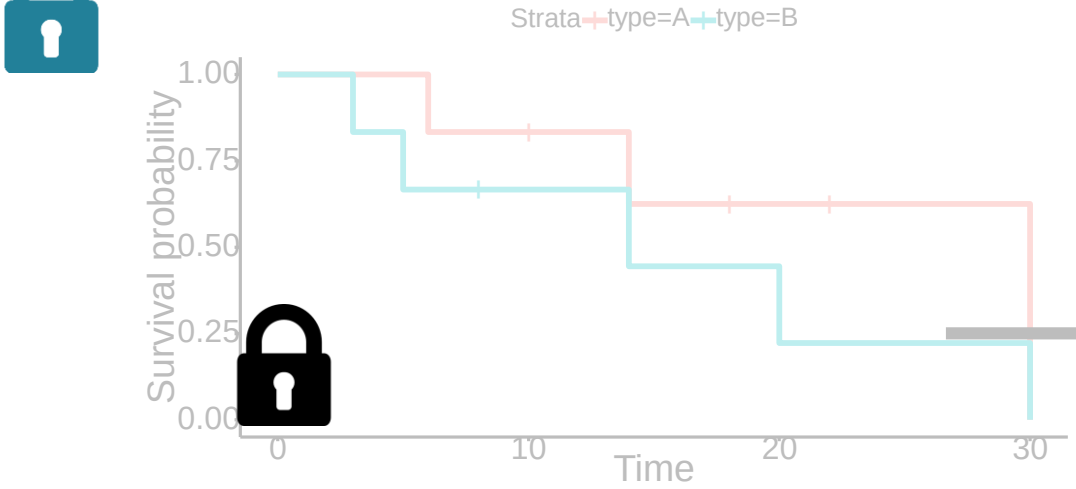


# Privacy-preserving Kaplan-Meier

- Non-parametric method to estimate survival curves

Time of death or time of censoring known per patient

Patient type

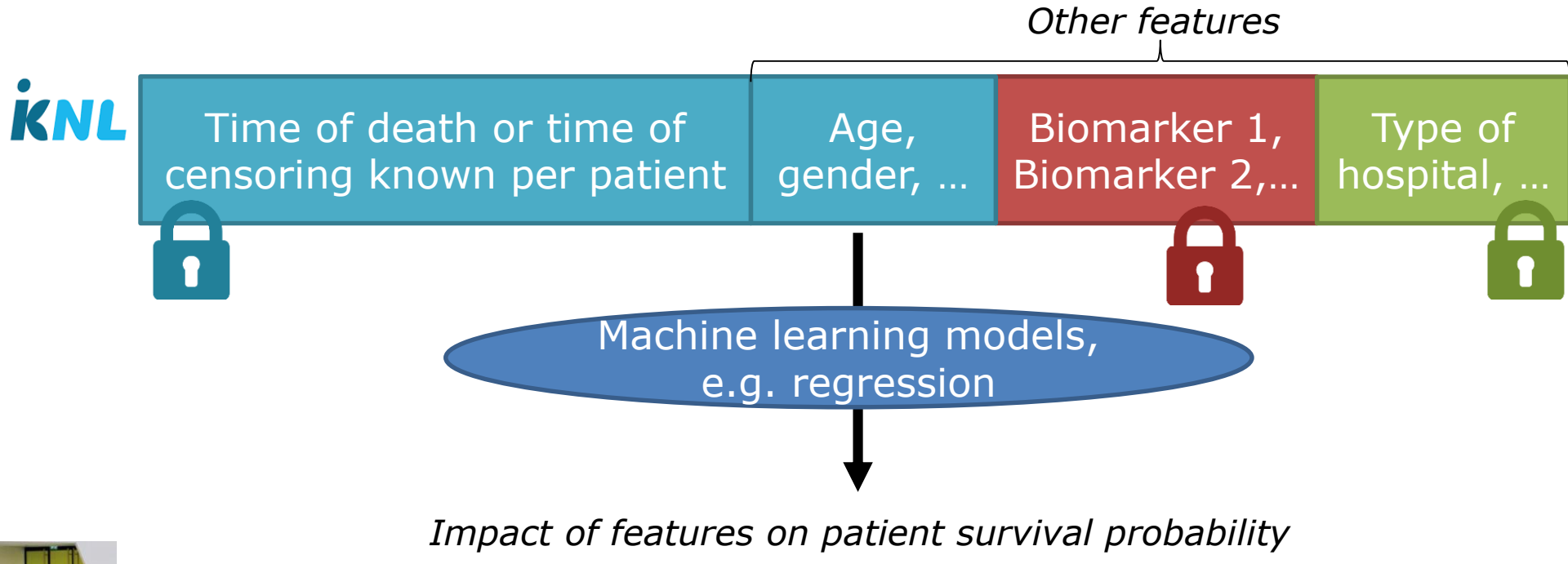



Statistical difference of survival rates between types

**Compute on encrypted data to obtain output of statistical test**

**Survival rates per type computed using Homomorphic Encryption *But not revealed!***

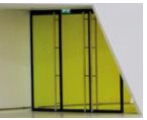
# Research into more complex survival analyses



# Conclusions

- Innovative technologies enable analyses on federated patient data that were impossible before
  - Sensitive patient data is cryptographically protected from all parties

- IKNL and TNO collaborate to bring these techniques to practical applications



# Contacts


- Daniël Worm
- daniel.worm@tno.nl
- www.tno.nl/mpc

- Gijs Geleijnse
- G.Geleijnse@iknl.nl
- www.distributedlearning.ai