

C4DHI Anniversary Workshop, June 3 2024, Prague

Artificial Intelligence for SSH

To Teach or Not To Teach?

Barbora Hladká

hladka@ufal.mff.cuni.cz



Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics



unless otherwise stated

Starting point

- Intro to Machine learning in R system ([url](#)) by Martin Holub&Barbora Hladká@[IFAL](#)
 - Term course at the Faculty of Mathematics and Physics, CUNI
 - Theoretical background and practical algorithms of Machine Learning
 - Programming in R system
 - 90 min lecture + 90 min lab session a week
 - Retired in 2022/23 after 10+ years
- Idea
 - Redesign the course for students of Social Sciences and Humanities (SSH)

Why to teach... 1. Data



These all are data.

Data

= information in digital form for computer processing

- text, audio, video, image, software
- born-digital = originate in a digital form
 - e.g. e-books, digital sound, video recordings
- digital reformatting = analog → digital
 - e.g. scanning books

Why to teach... 2. Data processing



These all are data.

Automatic data processing

- It's impossible to handle large data manually
- Integration of diverse data sources
- Reproducibility and transparency

Why to teach... 3. Future jobs



These all are data.

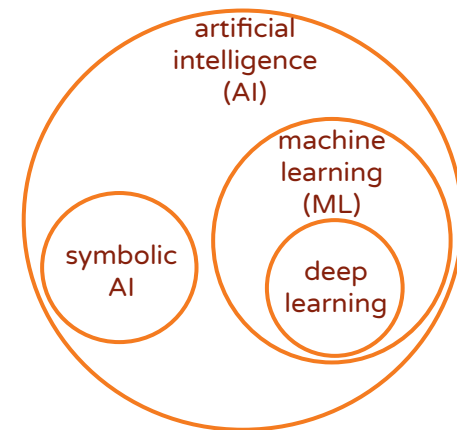
- Rapid advancements in ML/AI
- New career opportunities
- Equip students with the skills necessary to face new challenges

- Artificial Intelligence is a field of Computer Science simulates human intelligence processes by computer systems ([Wikipedia](#))
 - popularized in movies and literature for decades
 - associated with futuristic visions of humanoid robots, self-driving cars, superintelligent systems,...
 - high public interest and media coverage
 - > making it a buzzword
- Machine Learning is a field of Artificial Intelligence enables computers to learn from data ([Wikipedia](#))
 - technical understanding does not have the same level of excitement as more futuristic AI scenarios



Finding

Although the course is on ML, use AI in the course title and annotation.



Findings – CS and SSH students

Their motivation to enroll in a course on AI/ML

CS – Mostly self-motivated

- Intellectual curiosity
- Desire to contribute to the development of cutting-edge technologies
- Research interests
- IT job market

SSH – Explicit motivation needed

- What will I gain from this?
- Findings
 - Assist in bridging the gap between their fields and the technical nature of ML
 - Highlight the interdisciplinary nature of ML
 - Help overcome barriers to programming

Findings – CS and SSH students

Their background knowledge

CS

- Problem-solving abilities
- Strong foundation in mathematics
- Proficient in programming languages
- Familiar with algorithms and data structures

SSH

- Skilled in critical thinking, qualitative analysis, and writing
- Domain knowledge to interpret data
- Limited mathematical background, primarily basic statistics
- Different levels of programming knowledge, often minimal

New course for SSH students – Learning objectives

- Basic understanding of ML concepts
- Ability to use ML tools in SSH fields
- Emphasis on interdisciplinary applications

Compare them with the goals for CS students

- In-depth understanding of ML algorithms and their implementation
- Ability to optimize and troubleshoot complex ML models
- Focus on technical aspects and application in technology-driven fields

New course for SSH students – Syllabus

Part I – Prelude to Artificial Intelligence methods

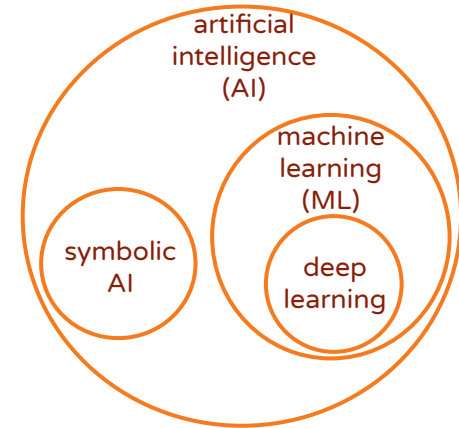
- General principles of Artificial Intelligence and statistical Machine Learning
- Historical overview of Artificial Intelligence from technological and user perspectives
- Statistical data analysis
- Technologies for Natural Language Processing
- Tidyverse Package in R

Part II – Traditional methods of statistical Machine Learning

- Supervised learning, classification and regression
- Application of selected learning algorithms
- Clustering
- Experiment evaluation

Part III – Deep Learning and Neural Networks

- Neural Network Architecture
- Representation of textual data using embeddings
- Training Neural Networks



New course for SSH students – Methodology

- Intuitive explanation of basic concepts instead of heavy mathematical formulations
- Case studies from SSH
 - Textual data only
 - Research questions
 - Teaching by examples, hands-on experience in R system



New course for SSH students – Methodology

dataset	topic	domain	source
Titanic	analysis + bin. classification	sociology	Kaggle
Migrants	analysis + NLP + classification	sociology	LINDAT/CLARIAH-CZ
Andersen	analysis + NLP	literary science	SMLTAR
Authorship	analysis + NLP + classification + NN	literary science	Kurz NPFL054
Horses	NLP + classification	linguistics, history	ZVařáková
Court	NLP + lin. regression + clusters + NN	law	SMLTAR
Campaigns	NLP + NN bin. classification	business	SMLTAR
Real.Estate	regression + NN	real estate market	DLR
Přijímačky	bin. classification	education	PAT
Height	correlation + clustering	psychology	PAT

New course for SSH students – Organization

- 90 min lecture + 90 min lab session (tutorial) a week, 14 weeks in total
- Lab sessions
 - Cloud computing Jupyter notebook, RStudio plugin
- Grading
 - Class participation is not mandatory
 - But we strongly recommend to attend all classes since the course is fairly intense
 - All homework assignments must be submitted
 - In-class final test and exam
- Supplementary course
 - For those who do not have sufficient background in mathematics and R programming
 - 90 min a week

New courses for SSH students – Implementation

Launch in 2024/25

- [Artificial Intelligence for Humanities](#)
- [Data Processing and Analysis for the Humanities](#)



New course for SSH students – Implementation

Test run of AI for Humanities in 2023/24 ([url](#))

- In cooperation with C4DHI@FSS CUNI
- 5 weeks
- Topics from Part I and II
- 90 min lectures, 90 min lab sessions, 4 HWs, final test, workshop on data oriented projects
- Teaching materials posted on the course website prior each class
- Without the supplementary course



New course for SSH students – Implementation

Test run of AI for Humanities in 2023/24 ([url](#))

Lessons learned

- The essence of students' work is fundamentally different from what they do in their SSH study programmes
- Great amount of work in classes and at home needed
- Undergrads require more explicit motivation than grads
- Terminology setting needed
- Notable progress after 5 weeks



Teaching by examples – Illustration

Research question: What influenced who survived the Titanic disaster?

```
library(tidyverse)
dataset <- read_csv("titanic.csv")
print(dataset)      # rows=passengers, columns=attributes
```

```
# A tibble: 891 × 12
  PassengerId Survived Pclass Name          Sex    Age SibSp Parch Ticket   Fare Cabin Embarked
  <dbl>      <dbl> <dbl> <chr>      <chr> <dbl> <dbl> <dbl> <chr> <dbl> <chr> <chr>
1         1         0     3 Braund, Mr. Owen Harris male    22     1     0 A/5 2...  7.25 NA     S
2         2         1     1 Cumings, Mrs. John Bradl... fema... 38     1     0 PC 17... 71.3  C85     C
3         3         1     3 Heikkinen, Miss. Laina fema... 26     0     0 STON/...  7.92 NA     S
4         4         1     1 Futrelle, Mrs. Jacques H... fema... 35     1     0 113803  53.1  C123    S
5         5         0     3 Allen, Mr. William Henry male    35     0     0 373450   8.05 NA     S
6         6         0     3 Moran, Mr. James male    NA     0     0 330877   8.46 NA     Q
7         7         0     1 McCarthy, Mr. Timothy J male    54     0     0 17463   51.9  E46     S
8         8         0     3 Palsson, Master. Gosta L... male     2     3     1 349909  21.1  NA     S
9         9         1     3 Johnson, Mrs. Oscar W (E... fema... 27     0     2 347742  11.1  NA     S
10        10         1     2 Nasser, Mrs. Nicholas (A... fema... 14     1     0 237736  30.1  NA     C
# i 881 more rows
```

Teaching by examples – Illustration

```
survived.sex <-  
  table(dataset$Survived, dataset$Sex)
```

```
survived.sex # contingency table
```

	female	male
0	81	468
1	233	109

```
barplot(survived.sex,  
  main = "Survived and Sex",  
  xlab = "Sex",  
  ylab = "Passenger count",  
  col = c("blue", "red"),  
  legend.text = TRUE,  
  args.legend = list(x = "topleft")  
)
```

