CHANGING WORLD OF ARTIFICIAL INTELLIGENCE, THE RECENT EXPLOSION

Ivan Bratko
University of Ljubljana
Faculty of Computer and Information Science

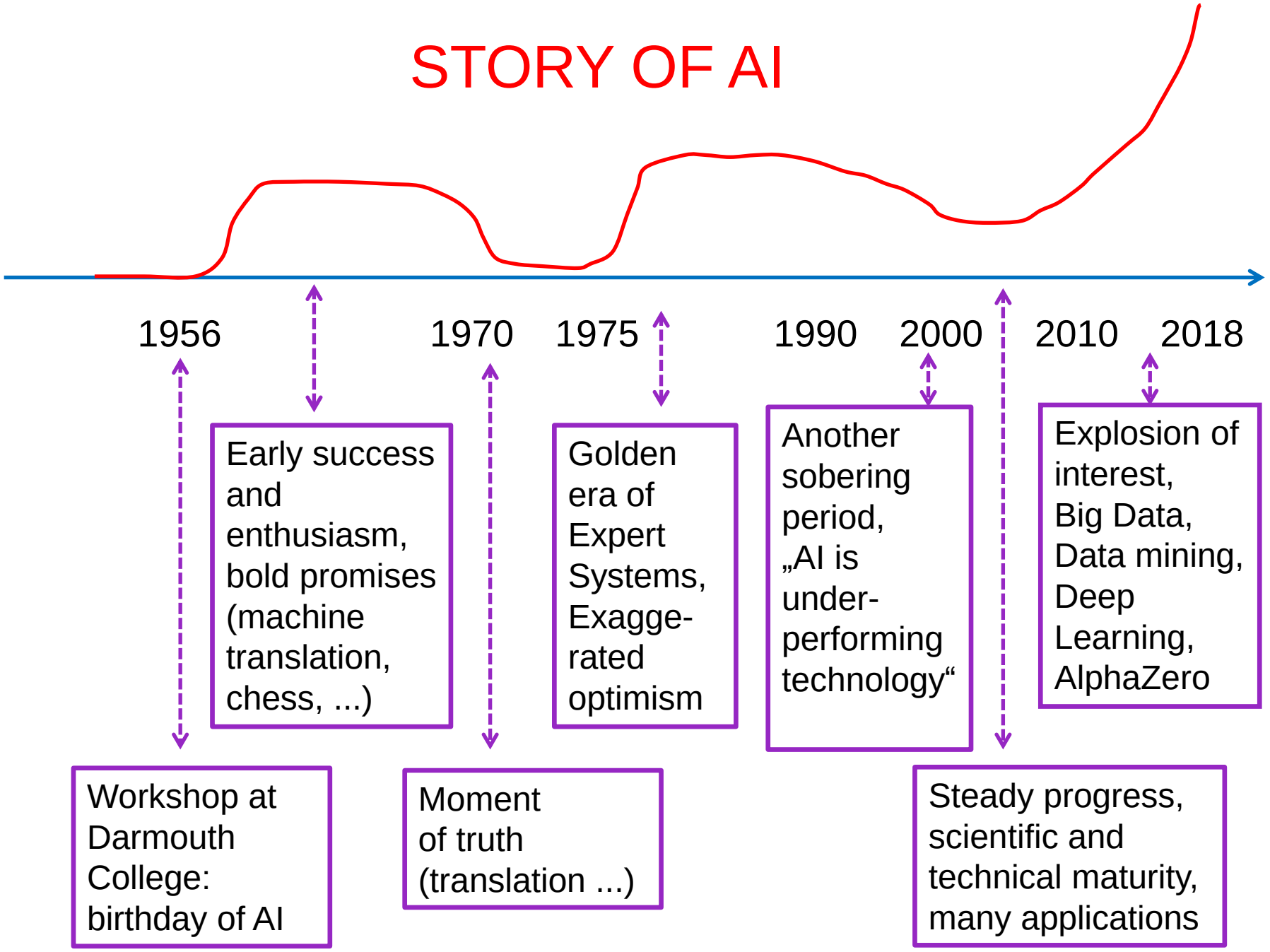
CONTENTS

- Brief history of Artificial Intelligence
- Current trends in technical areas of AI
- Machine learning and breakthrough in Deep learning
- AI applications: ethical and legal issues

BRIEF STORY OF AI

There were some
UPS and DOWNS
before it truly took off

STORY OF AI


1956
1970
1975
1990
2000
2010
2018
Workshop at Darmouth College: birthday of AI
Early success and enthusiasm, bold promises (machine translation, chess, ...)
Moment of truth (translation ...)
Golden era of Expert Systems, Exagge-rated optimism
Another sobering period, „AI is under-performing technology“
Steady progress, scientific and technical maturity, many applications
Explosion of interest, Big Data, Data mining, Deep Learning, AlphaZero

TYPICAL PROBLEMS FOR AI: FROM BEGINNING TILL NOW

FROM TOY PROBLEMS TO VERY RELEVANT, HARD PROBLEMS

Artificial Intelligence 1960,

Carnegie-Mellon Uni., Pittsburgh

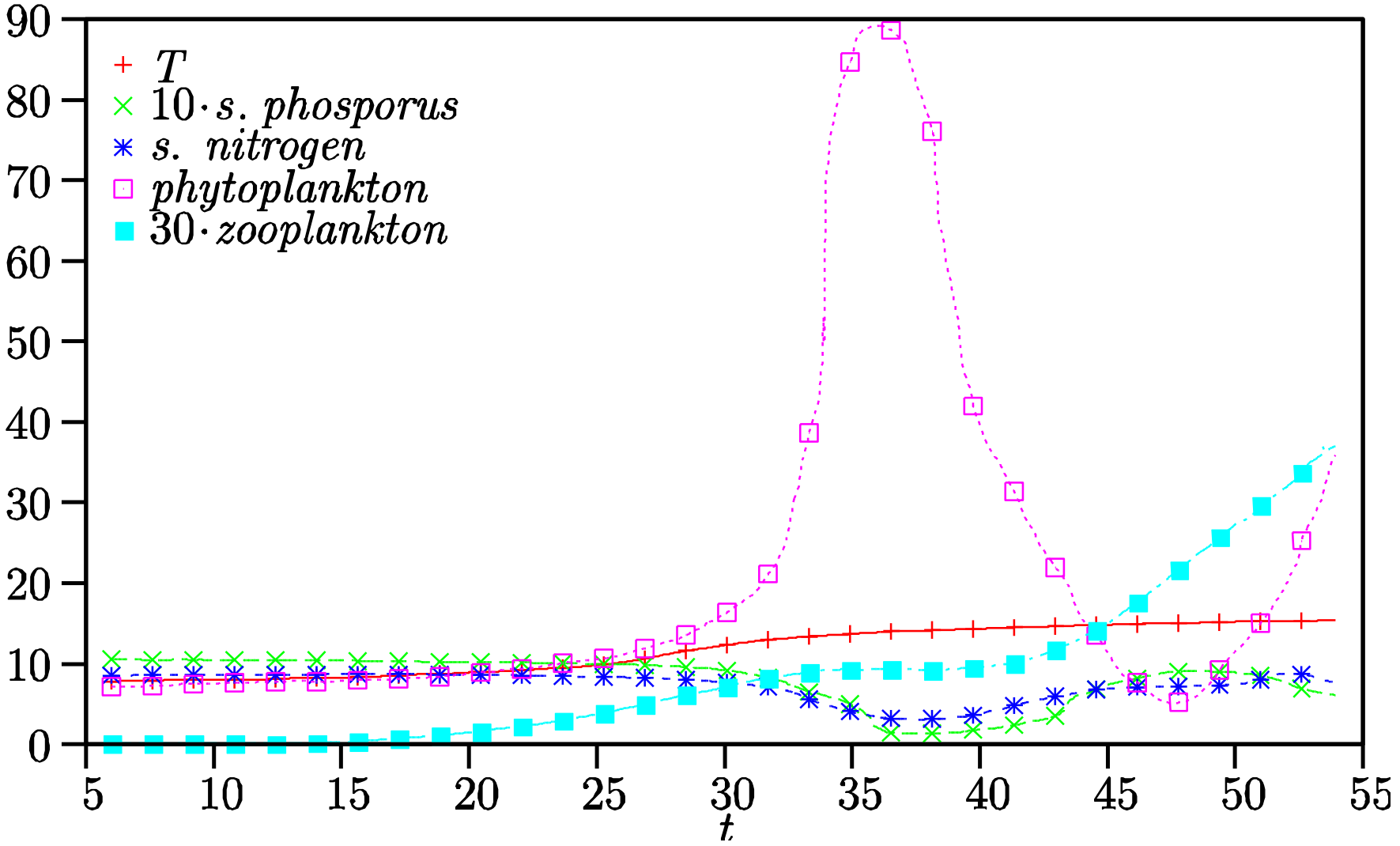
Cryptarithmetic puzzles

```
   D O N A L D
+  G E R A L D
-------------
=  R O B E R T
```

How humans solve such problems? (Simon and Newell)

What algorithms can simulate human problem solving?

Mid 1980s, Danish lake Glumso



Computer to find laws of algae growth

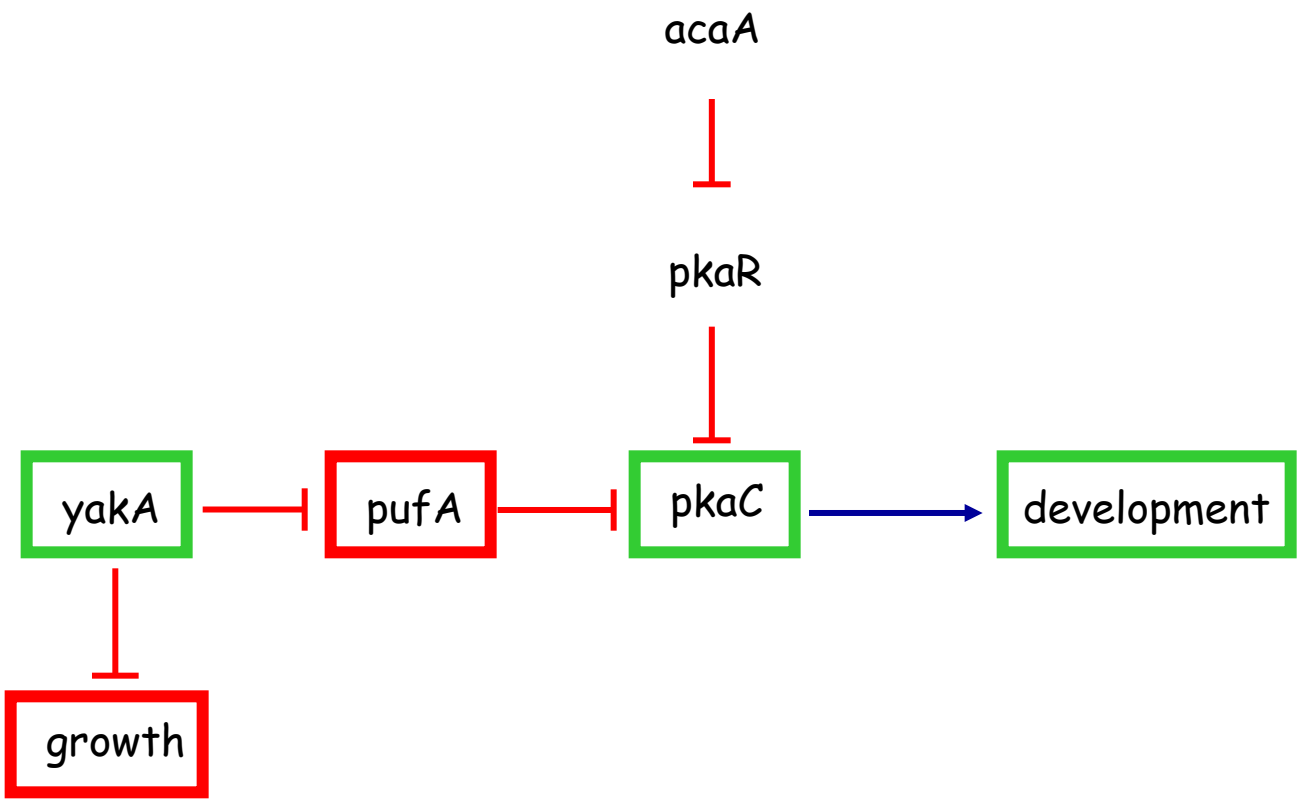
2000s, Ameba *Dictyostelium*

When food is cleaned, *Dictyostelium* get together and converge in mound.

Development: the mounds stretch into slugs, which topple over and crawl away.

Computer to discover a theory of how genes effect functions of the organism?

Genetic theory discovered by GENEHYP



How can computer discover such a theory? For just 7 genes there are of the order ***1.000.000.000.000.000*** possible networks! (Zupan et al. 2001)

TYPICAL 2010s APPLICATION AREAS

- Recommender systems
- Machine translation
- Image and speech recognition
- Intelligent sensors
- Finance

SOME HIGHLIGHTS 2010s

- Self-driving cars
- Intelligent robots
- Automatically generating legislation for humans (traffic regulations, IJCAI'11 best video)
- Helping humans to develop their creative style for music composition and story writing (IJCAI'13 best video)
- Automatically constructed world-wide map of wealth/poverty from satellite images, no survey data (Ermon, IJCAI'18, Computers and thought award)
- AI in manipulating human opinion through Internet

WHAT USED TO BE DIFFICULT FOR AI HAS CHANGED

- For example, consider computer chess or Go:
  - Once: How to play well and beat human players?
  - Now: Performing the task is not so hard any more.
  - What is hard now is: How to **explain** solutions to humans, how to **comment**, how to **teach** humans?
- A chess program may play a perfect game, but cannot explain it in human-understandable terms
- New area: “Explainable AI” (IJCAI’17), Intelligible AI

TECHNICAL AREAS OF AI, VERY ROUGHLY

- Problem solving and search
- Knowledge representation and reasoning
- Machine learning
- Natural language processing
- Perception, vision
- Robotics, robot vision, robot planning

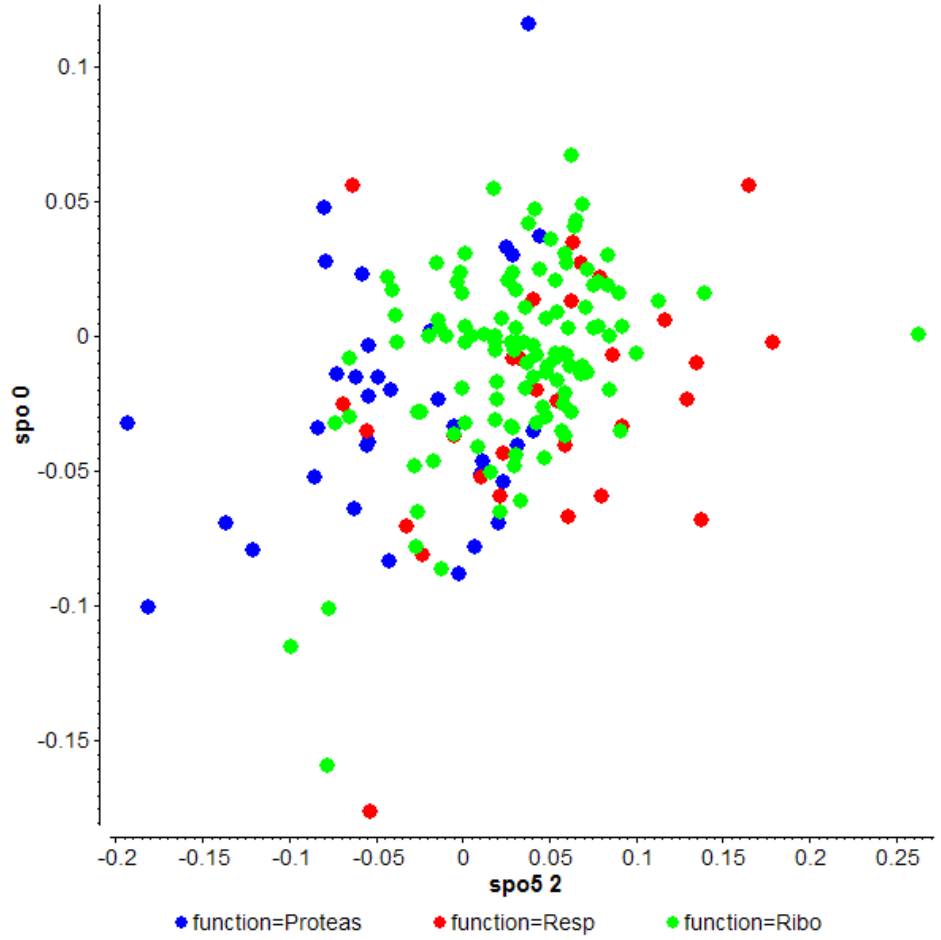
RECENT TECHNICAL PROGRESS IN AI

- Most significant technical progress in past 30 years has been in area of Machine Learning (ML)
- ML has been used in many successful applications
- ML has become standard tool in many disciplines in application and scientific research
- ML is now often essentially referred to by popular terms like: Data Mining, Big Data, Data Science
- Popular misconception today: AI = ML

EXAMPLE APPLICATION OF MACHINE LEARNING

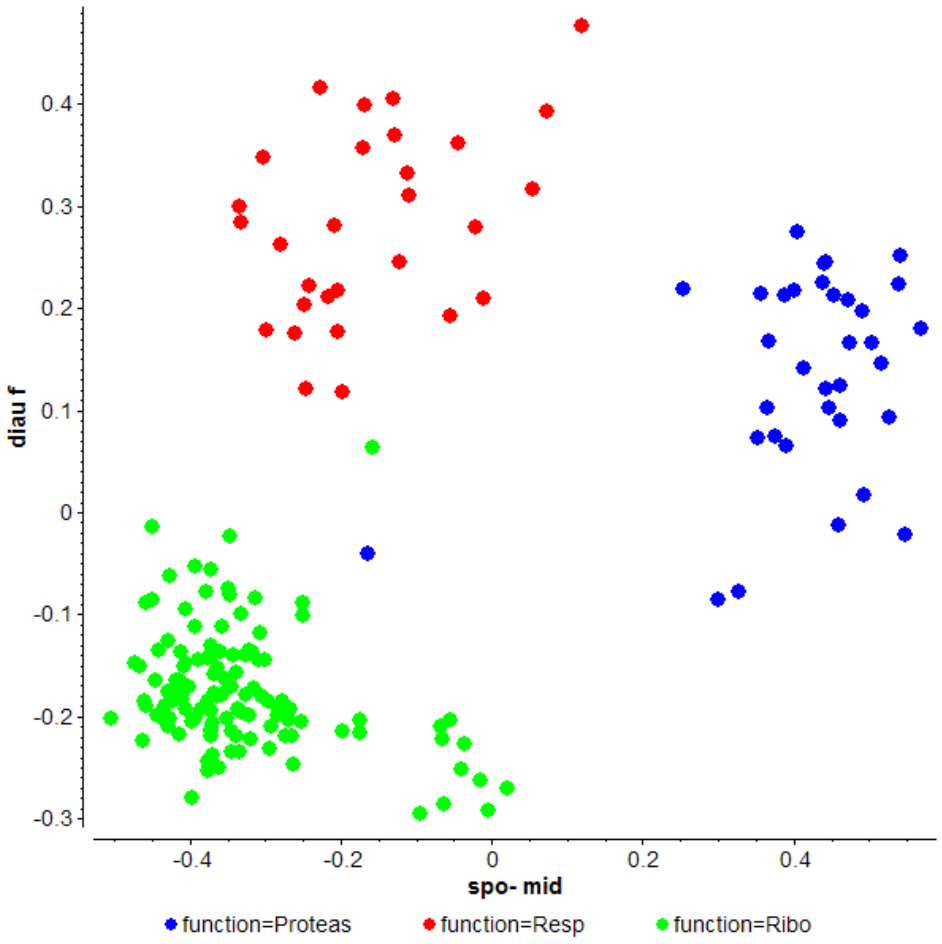
- Scientific discovery of laws in experimental data
- Using Machine Learning for intelligent data visualization
- Machine Learning tool used: ORANGE learning system (Zupan, Demšar et al.)

Brown data set – functions of genes, “random” vizualisation



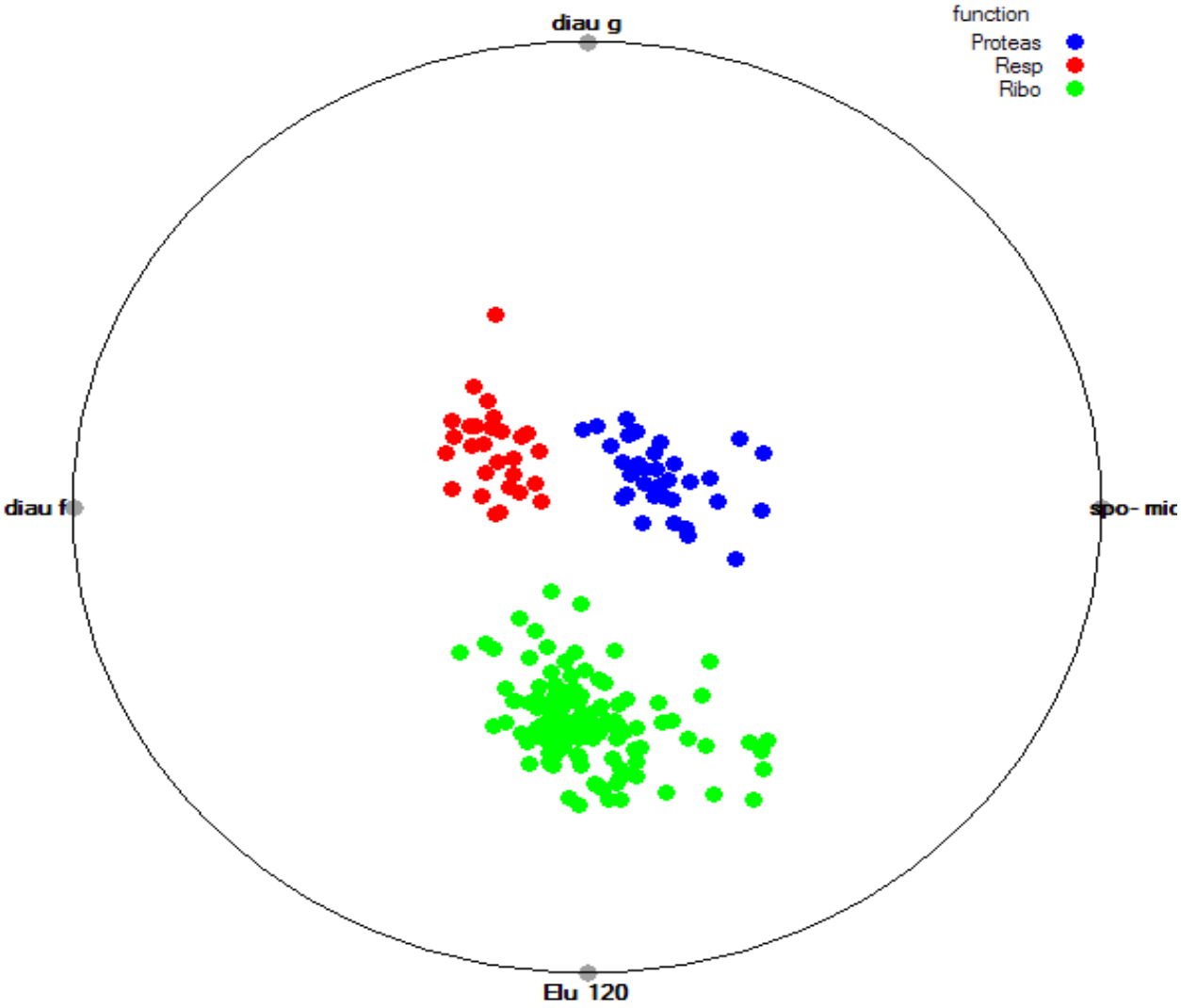
There seems to be no regularity, all colours mixed up

Brown data set – Intelligent visualization, Vizrank



Suddenly everything seems to be clear, all colours nicely separated

Brown data; Intelligent visualisation with Vizrank Radviz



THE RISE OF DEEP LEARNING

- Deep Learning is an approach to ML based on Artificial Neural Networks (ANN)
- Deep neural networks are networks with many “hidden” layers
- This is *the* approach to ML with the most spectacular improvements in 2010s

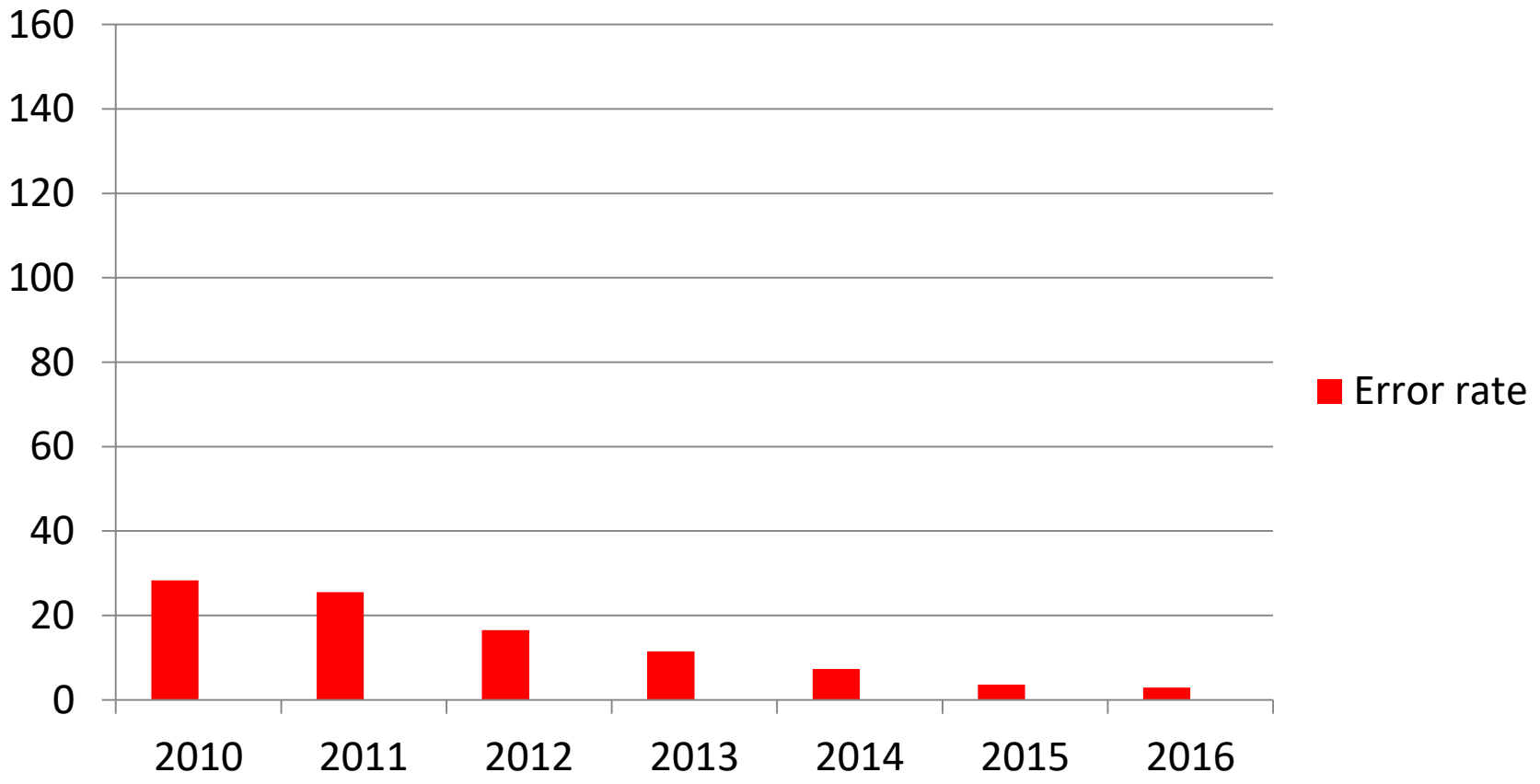
Yann LeCun talk on Deep Learning at IJCAI 2018, Stockholm

- For speech recognition, image recognition, face recognition, generating captions for photos, machine translation, ...

- Traditional ML: user has to define useful features in image

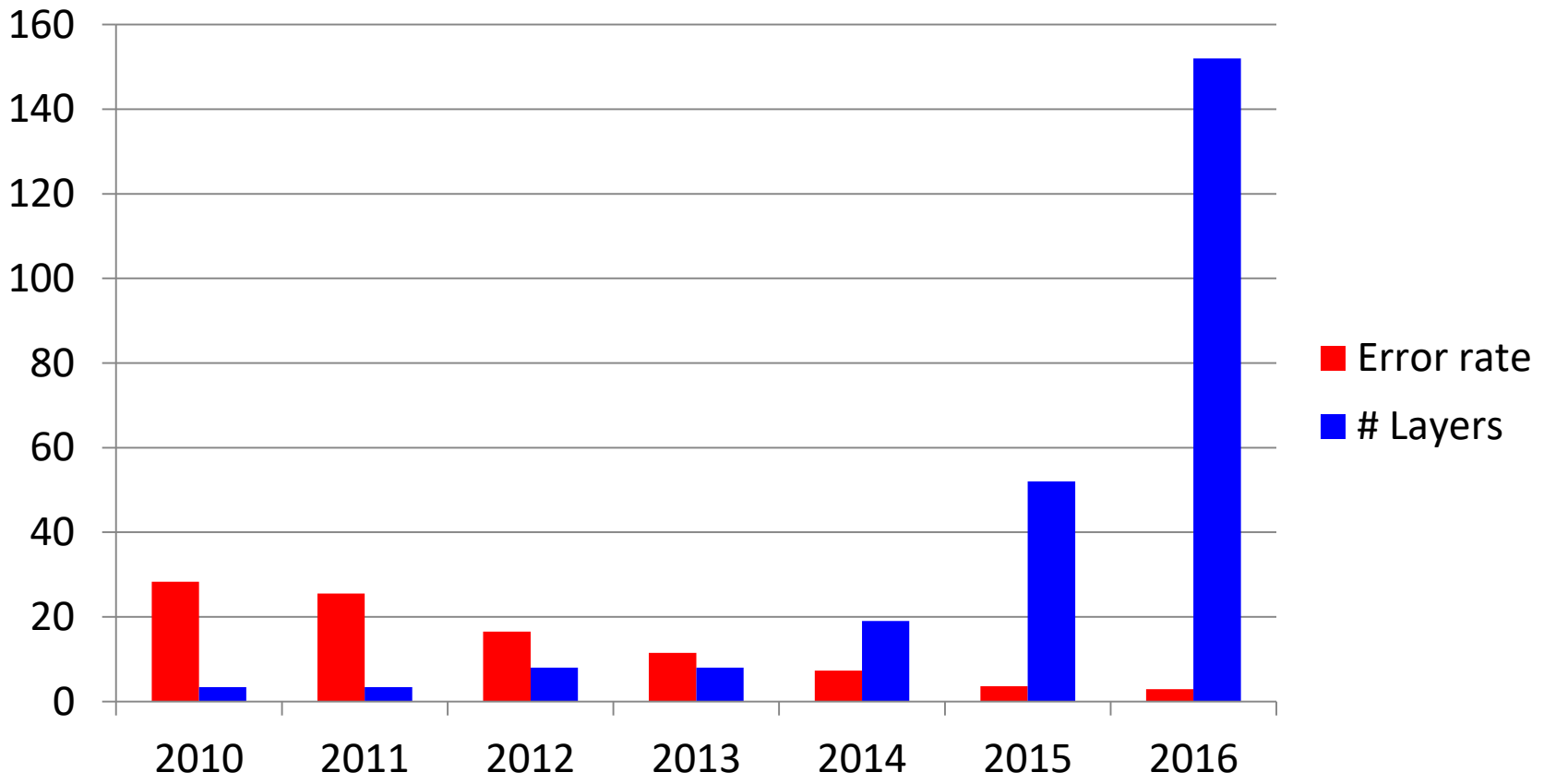
- Deep learning constructs good features automatically: you can feed raw image into network, and it will construct features on its own

PROGRESS OF DNNs ON ImageNet DATASET 2010-2016



From LeCun 2018

PROGRESS OF DNNs ON ImageNet DATASET 2010-2016



Learning times:
2010: ~1 month; 2016: ~1h on GPU network

From LeCun 2018

WHY DO WE NEED SO MANY LAYERS?

- Probable answer: natural data are compositional
- Images are made of low level to high level concepts: from pixels to edges to contours to ... to parts to objects
- By learning in multiple layers, a DNN automatically builds a representation that can be used in a different domain (learning a hierarchy of useful concepts in the domain)
- However, it is rather difficult to understand exactly what the learned representation is

TYPICAL APPLICATIONS OF DEEP LEARNING

- Image classification, recognition (medical images, faces, driving vehicles)
- Language translation
- Content understanding (search, filtering, ranking)
- Science (biology, genomics, physics)
- Games
- Virtual assistants – not so good!

FAMOUS EXAMPLE OF FOOLING A DNN
From Goodfellow et al. 2015

- DNN learns to recognise pictures of panda

FAMOUS EXAMPLE OF FOOLING A DNN

- What is this?

NN says: This is gibbon, 99% confidence

FAMOUS EXAMPLE OF FOOLING A DNN, HOW COULD THAT HAPPEN?

- fix trained network
- carry out backprop using *wrong* class label
- update input pixels:

Goodfellow, Shlens, and Szegedy, ICLR 2015

SPECTACULAR SUCCESS OF Alpha/X PROGRAMS FROM DeepMind

- 2016: AlphaGo (Silver, ..., Demis Hassabis, Nature 2016)
  AlphaGo defeated a leading Go-player Lee Sedol
  (first time that a computer performed better
  than best human Go player)

- AlphaGo learned from examples of good games
  played by humans

- October 2017: AlphaGo Zero (Silver, ..., Hassabis, Nature)
  AlphaGo Zero vs. AlphaGo 100:0
  AlphaGo Zero only learned by self-play (no access to human knowledge about the game; only the rules of the game)

AlphaZero

- December 2017: AlphaZero (Silver, ..., Hassabis, 2017)
  In less than 24 hours of learning by self-play it became better than any human and any other program at chess, Go and shogi

- In 90 min of self-play, AlphaZero outperformed best humans chess players

TECHNIQUES USED BY AlphaZero

- Uses Reinforcement Learning to learn by self-play (simulated games against itself)
- Uses Deep Neural Networks to generalize results of RL: learn to predict values and move probabilities for positions never encountered before
- Uses Monte Carlo Tree Search to select the move to play

ARE THERE ANY LIMITATIONS TO AlphaZero APPROACH?

- Limitations:
  - Relies on virtually unlimited self-play, i.e. unlimited simulation; in many real-world domains this is not possible (in medicine, nature produces new examples, not a simulator)
  - Inability to explain its decisions to human players (example games between AlphaZero and Stockfish)
- See: AlphaZero – What's Missing? (Bratko, 2018)

LIMITATIONS OF DEEP LEARNING

- LeCun: Why virtual assistants with DNNs do not work well?

  DNNs work well for recognition/perception,

  but they lack reasoning

- But, automated reasoning is a major traditional area of AI

- Obviously: DNNs should be combined with reasoning techniques - this may not be so straightforward

LECUN: LIMITATIONS OF DEEP LEARNING

- Deep learning needs “Big data”
- What if big data is not available? (Like in Medicine)
- Humans can still learn, even from very “small data”
- Key for humans: Use prior knowledge
- Can DNN be made to use prior knowledge?
- Note IB: There are other approaches to ML that can very naturally and flexibly use prior knowledge. For example Inductive Logic Programming (ILP)

AI technology enables applications that raise moral or ethical issues

SOME ETHICAL ISSUES IN AI APPLICATIONS

- Legal issues in self-driving cars:
  - Who is responsible? Can it be the car?
  - Crash unavoidable, how will car decide between alternatives who to kill: a little girl or an elderly couple (Bonefon's study 2018)?
- Lethal autonomous weapons: May decision to kill be delegated to a computer?
- Manipulation of people over social networks – end of democracy?

SCIENTIFIC AMERICAN, February 2017

Will Democracy Survive Big Data and Artificial Intelligence?

Dirk Helbing, Bruno S. Frey, Gerd Gigerenzer, Ernst Hafen, Michael Hagner, Yvonne Hofstetter, Jeroen van den Hoven, Roberto V. Zicari, Andrej Zwitter

Helbing:

“In the future, using sophisticated manipulation technologies, these [software] platforms will be able to steer us through entire courses of action ... from which corporations earn billions. *The trend goes from programming computers to programming people.*”

CAN AI TAKE OVER?

- Until recently, the answer was clear: No!
- Answer now is getting complicated: It’s becoming possible to see how this could happen ...

SOME INITIATIVES

- Future of Life Institute
  - Open letter of AI researchers 2015 (8,000 signatures)
  - Open letter against lethal autonomous weapons (26,000 signatures)
  - Konference Beneficial AI 2017, Asilomar AI principles (4,000 signatures)
  - Asilomar AI principles endorsed by State of California, 2018
- Responsible AI, Europen Network –
  status of proposal for two years already