ICIC Sitges 2007

Intelligent IT-Systems?

Challenges, Fakes and Hard Science

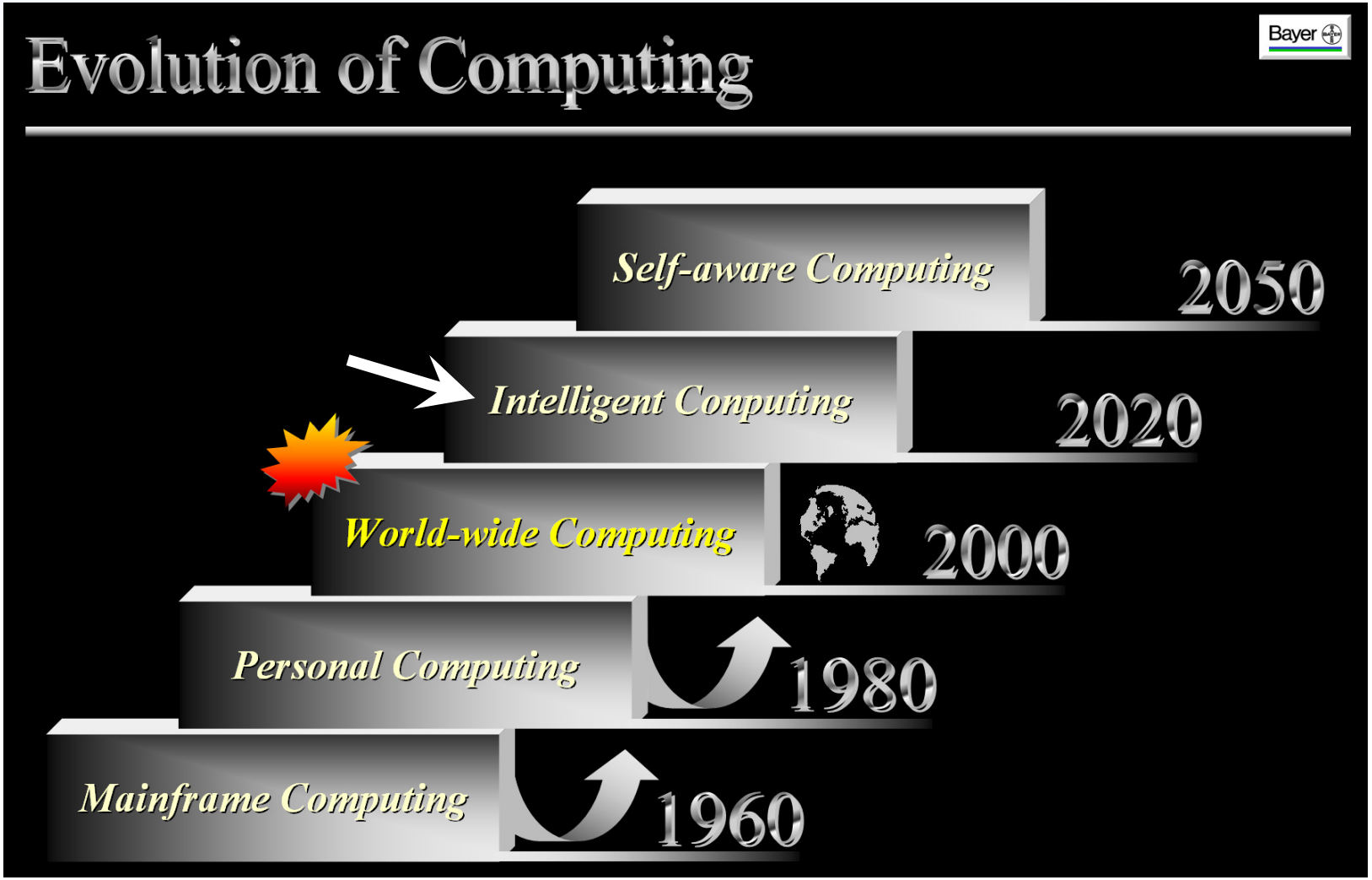
Stefan Neumann

GNWI – Gesellschaft für natur-wissenschaftliche Informatik

Achim Zielesny

University of Applied Sciences of Gelsenkirchen

ICIC Nimes 1998



Intelligent Computing

What means *intelligent*?

No precise definition!

Rule of thumb:

If a machine is capable of performing an *intelligent* operation this operation is **no longer** called *intelligent*.

Human Understanding

John went to a restaurant. He ordered lobster. He ate and left.

What did John eat?

Human answer: Lobster, of course!

Correct answer: I don't know!

Human Understanding

=

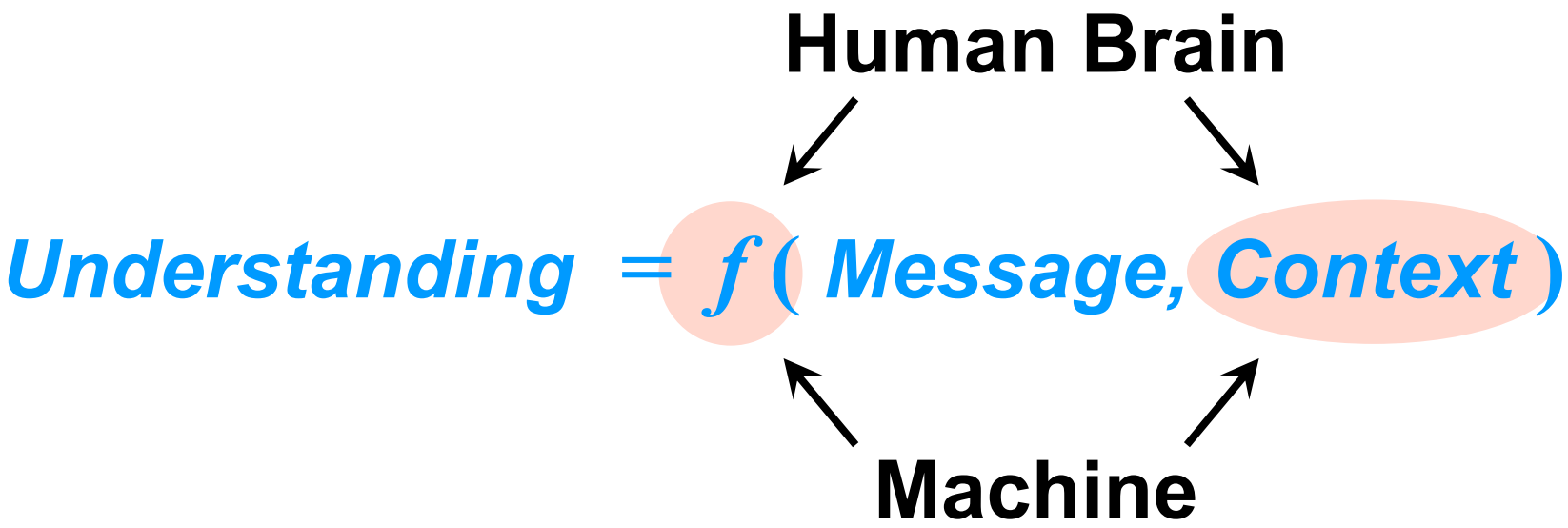
Interpretation within a learned cultural context

$$Understanding = f(Message, Context)$$

not separable

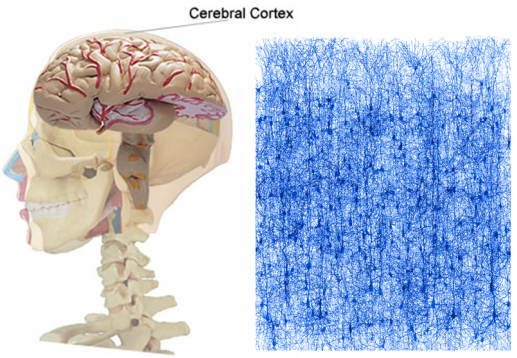
(True for Man **and** Machine)

Human Understanding



Complexity?

Complexity



Human brain: Most complex structure of the known universe

Dynamical system of …

- a **100 billion (10^{11})** asymmetrically interacting neurons
- with a **100 trillion (10^{14})** connections

Estimated calculation speed …

- **10^{16}** up to **10^{19} calculations per second**
- with a memory of **10^{12}** up to **10^{17} bytes** (**terabytes** up to **petabytes**)

Compare to this Notebook-Computer: 10^8 cps with 2 gigabyte memory

Complexity

Todays Machines << Human Brain

$$Understanding = f(Message, Context)$$

Numerous techniques: DB models based knowledge representation, enumeration of rules, heuristics, decision trees, machine learning/nonlinear mapping etc.

Fakes

(e.g. Joe Weizenbaum's ELIZA – simulates a human psychotherapist)

Fakes

- **Challenging**, **sophisticated**, **hard to make**
- May be very **successful** (comparable to human beings)
- May be **inspiring** and **helpful** (for new insights etc.)

But …

Example – Bioinformatics Project

Interdisciplinary mental disease R&D project

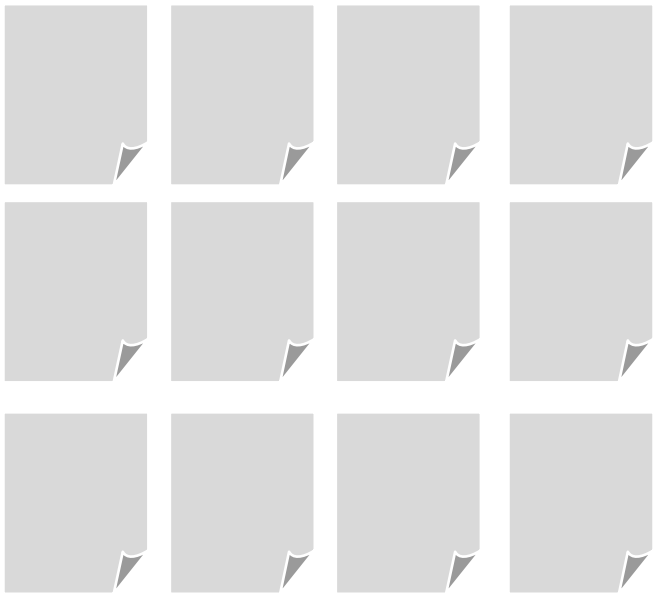
Btx goal:

Intelligent PubMed retrieval system for all project collaborators, i.e. a system that ***understands*** what the researcher wants

WYTIWYG – What you think is what you get

Example – Bioinformatics Project

Solution: 4-Step System



Step 1 – Basic

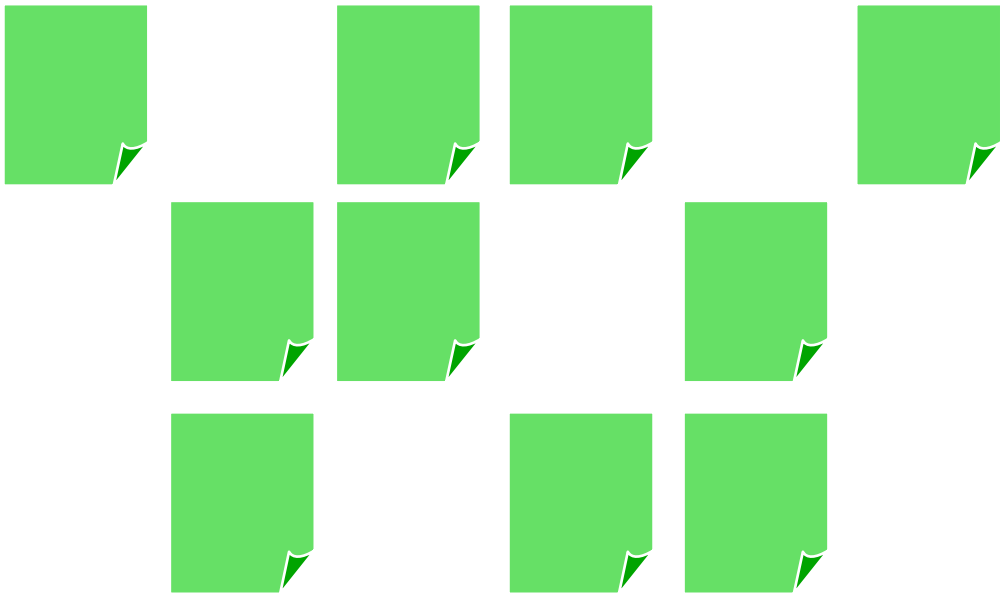
Example – Bioinformatics Project

Solution: 4-Step System

Step 2 – Enrichment

Example – Bioinformatics Project

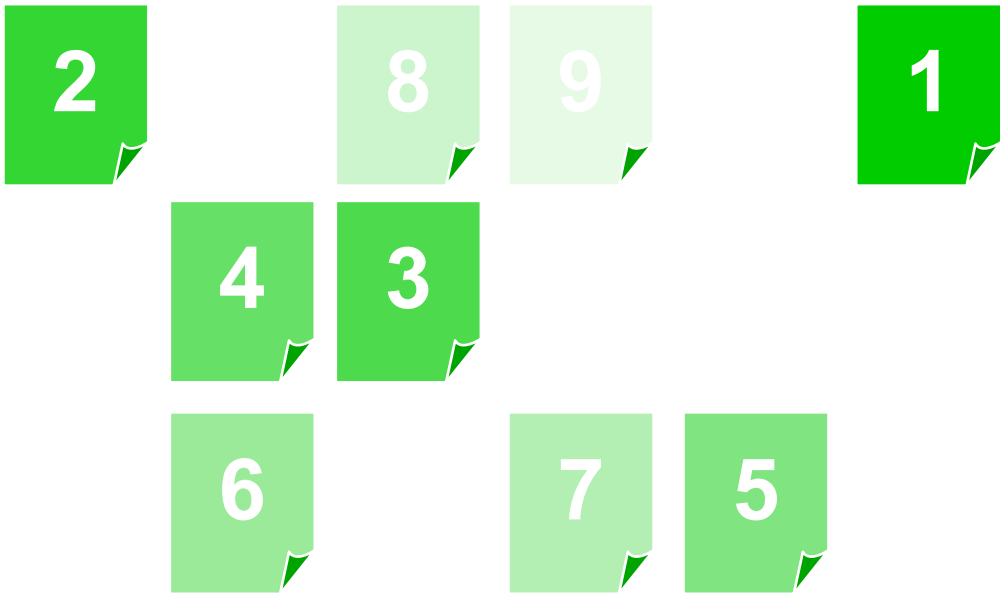
Solution: 4-Step System



Step 3 – Exclusion

Example – Bioinformatics Project

Solution: 4-Step System



Step 4 – Ranking

Example – Bioinformatics Project

User View:

Technical basis: Sophisticated data models for synonyms and semantic relations, thesaurus and ontology functions, enrichment/exclusion rules, decision trees, machine learning and nonlinear mapping etc. (“the full program”): **12 years of highly-qualified scientific work**

Result: Goal achieved

(**successful** system with **very high user satisfaction**)

But …

Example – Bioinformatics Project

- The system was very specific, i.e. restricted to a very narrow R&D field – and could **not be generalized**
- The system was **extremely hard** to manage and to keep up-to-date

Project end = System end

(“fire-and-forget software”)

Fakes are extremely expensive

(**a professional information broker achieves same/better results with <10% of the expenditure**)

Fakes – LL

- **Industrial Manager**:

  Don't spend money on fakes – but think about using publicly financed ones …

- **Public Manager**:

  Don't tell anybody what you have heard so far – keep on spending money …

Fakes are great for educational purposes

Alternatives?

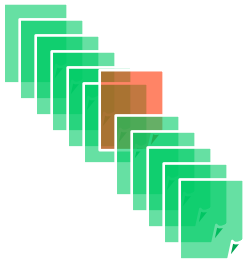
Is there anything better than fakes? *No*! (2007)

Alternative: Hard Science Solutions

Are based on already attributed human contexts
(they do **not** have to ***understand***)

Hard Science Solutions are **comparable** or **superior** to human beings (**on the same data**)

Example – Patent Similarity Ranking



International Patent Classification (IPC)

e.g. *A61K*-031

Section (*A*) + Class (61) + Subclass (*K*) + Group (031)

Depth:

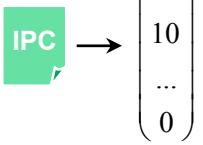
Section + *Class* + *Subclass*

e.g. *A61K*

	match score
Section	1
Class	3
Subclass	5
Group	7

$$\begin{pmatrix} A01N \\ A61K \\ A61P \\ \dots \\ C07H \end{pmatrix}$$

master vector

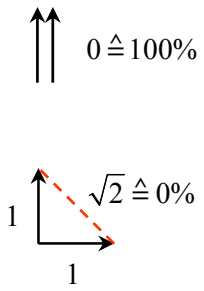


patent vector

Interesting patent | **IPC as descriptor** | **Set similarity depth** | **Set similarity score matrix** | **Create master vector (all IPCs)** | **Score of single patent IPCs**

$$\frac{1}{\sqrt{\sum_{i=1}^{L} x_i^2}} \begin{pmatrix} 2 \\ 6 \\ 10 \\ \dots \\ 0 \end{pmatrix}$$

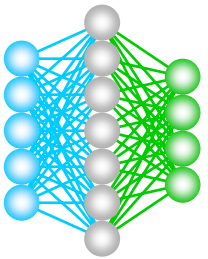
patent vector



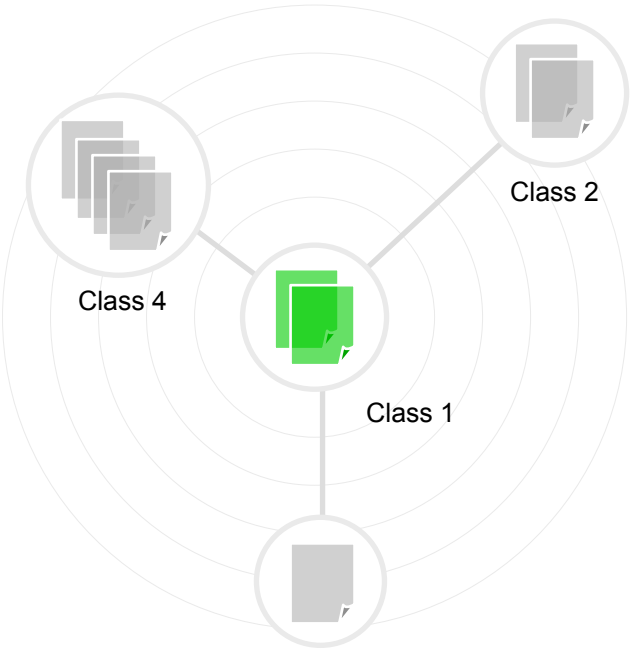
Rank#	Score	Documents
1	100%	
2	87%	
…		
8	0%	

Normalize patent vectors | **Calculate difference to _interesting_ patent** | **Result: Rank of similarity**

Example – Patent Similarity Classification



e.g. Neural Network



outer ring means maximum dissimilarity of classes, i.e. patents of class 1 and class 2 have *nothing in common*

Patent set → **Classifiy e.g. with similarity score** → **Inspect class similarities – select documents** → **Partial view**

Hard Science Solutions

- Relatively **wide range** of application
- **Fast**, **robust**, often **interpretable**
- **Easy** to implement (for professionals)

But …

Inferior to results of a good comparable fake!

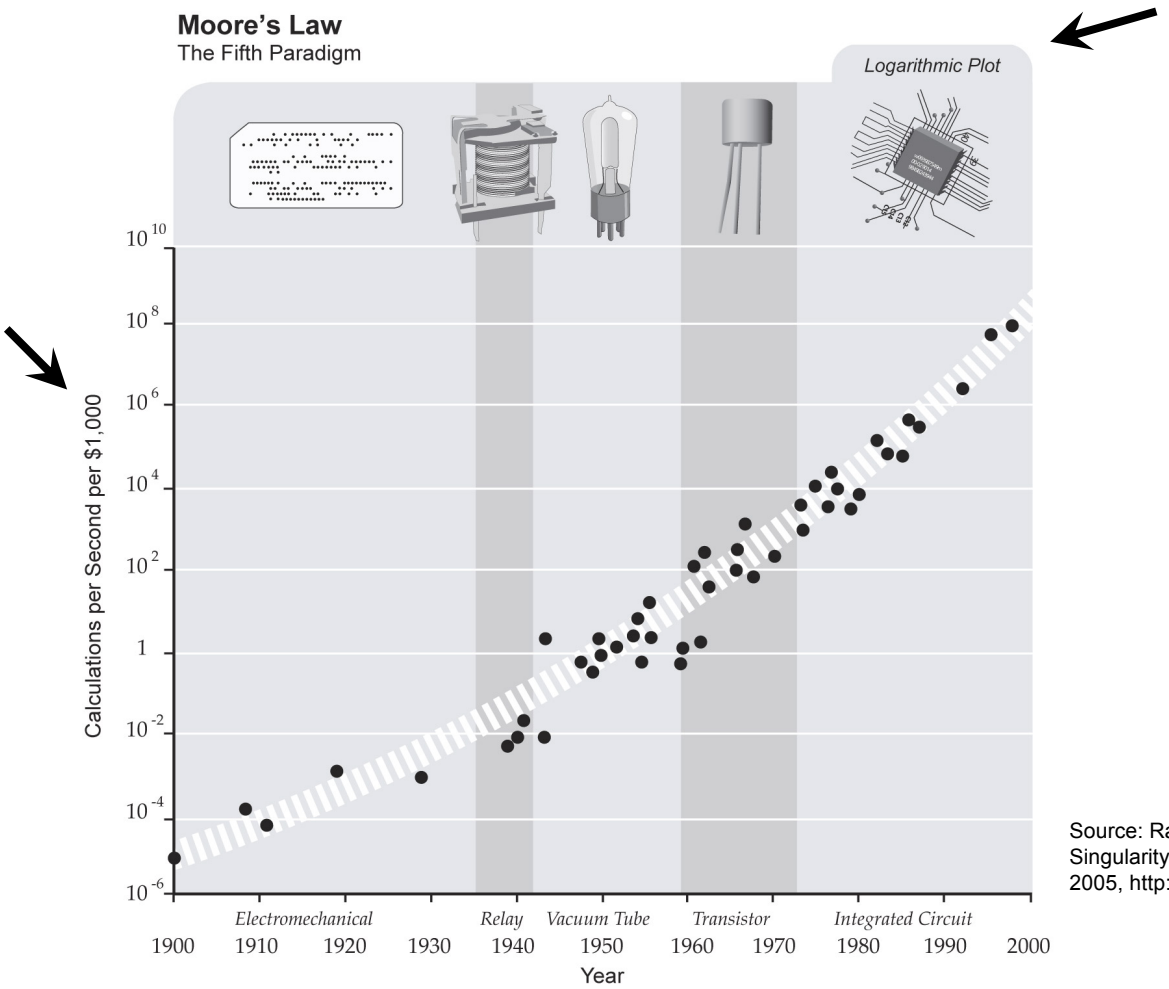
Hard Science Solutions – LL

Industrial & Public Manager:

Keep on spending money on **efficient technologies** for workflow optimization

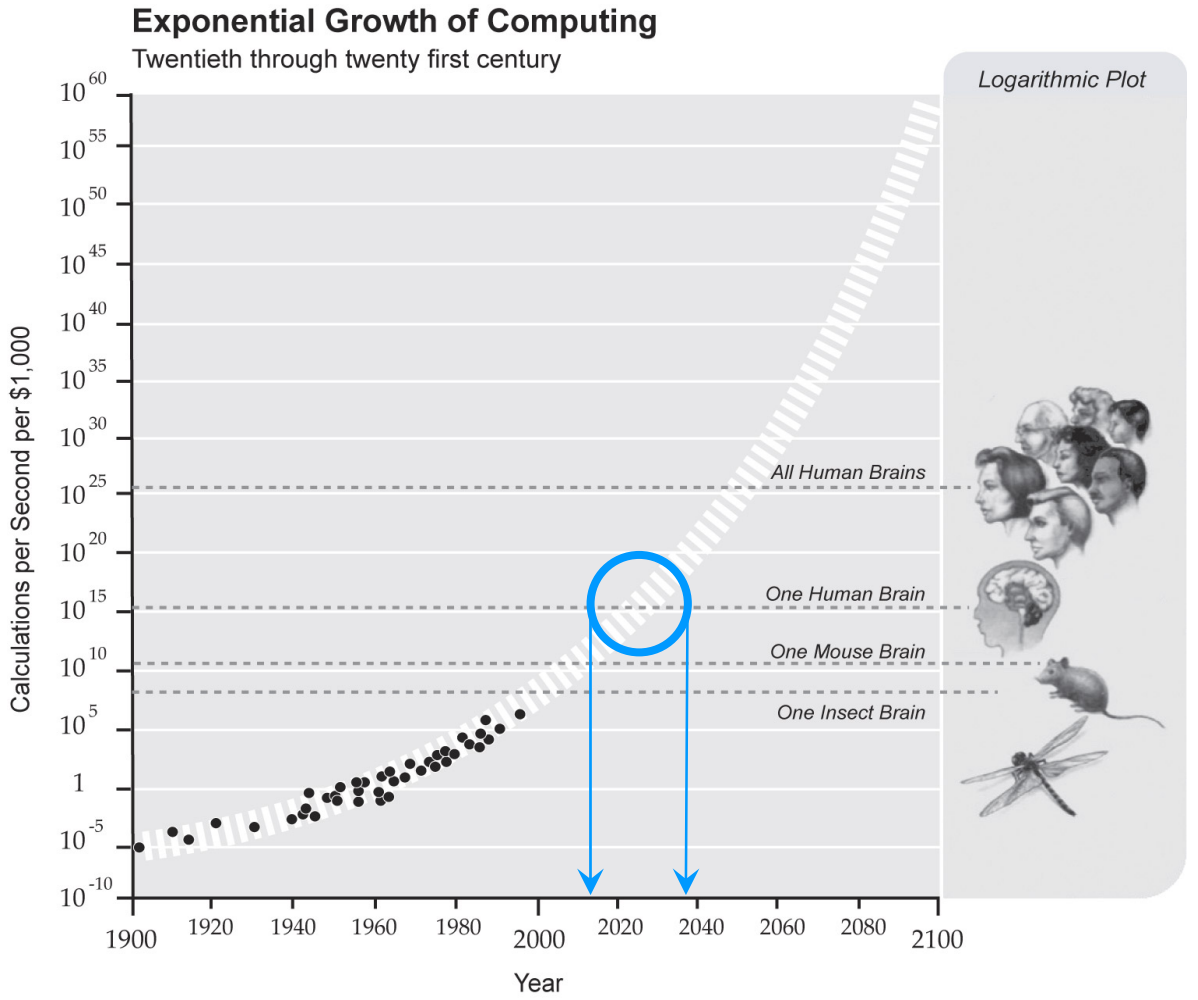
(Human beings can't do better and fakes will wrap them up …)

Complexity: The Power of Computing



Source: Ray Kurzweil, The Singularity Is Near, New York, 2005, http://singularity.com

Complexity: The Power of Computing



Source: Ray Kurzweil, The Singularity Is Near, New York, 2005, http://singularity.com