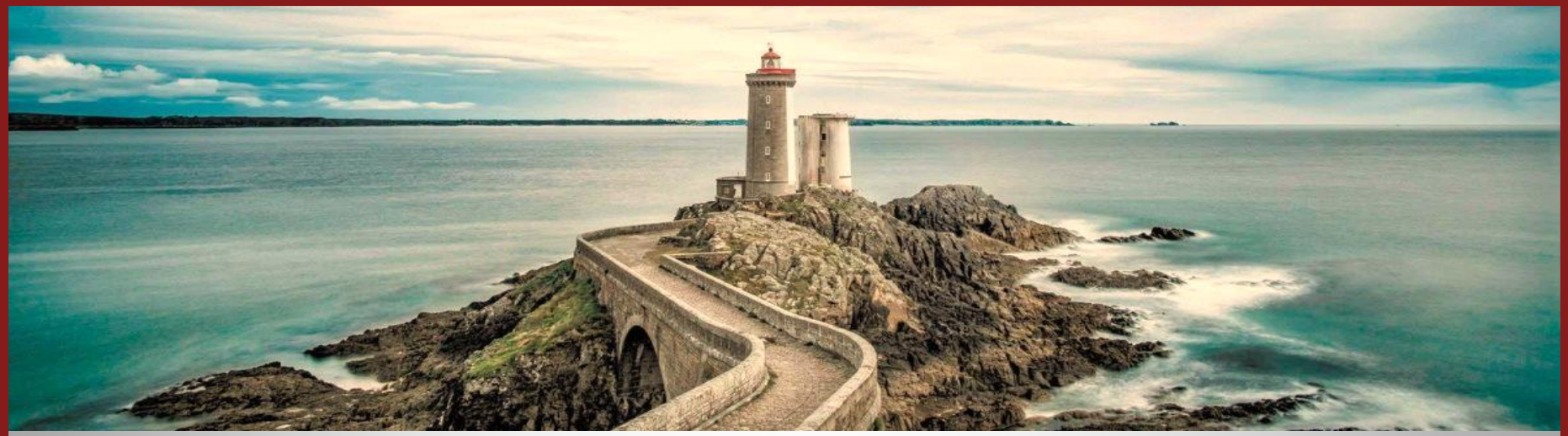


Course Review

上堂回顾

- 4.6 产权利益
- 4.5 真实可信
- 4.4 公平公正
- 4.3 隐私保护
- 4.2 自由尊重
- 4.1 健康安全

1. 关于机会: **重视技术的影响**
2. 关于工作: **社会学角度分析**
3. 谈数据利用: **需要慎重对待**
4. 谈以人为本: **值得反复强调**

Computer Ethics

计算机伦理学

十-5、真实可信

授课人: 李超 博士
chaol@sjtu.edu.cn

2020年 秋冬学期

上海交通大学计算机科学与工程系

Course Outline

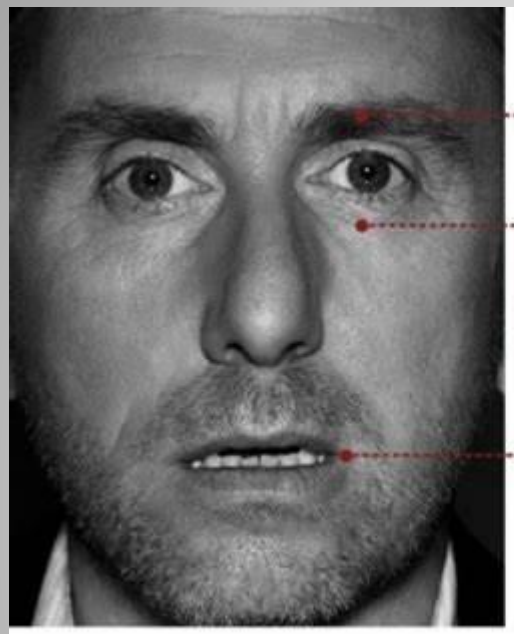
案例总览

- 4.6 产权利益
- 4.5 真实可信
- 4.4 公平公正
- 4.3 隐私保护
- 4.2 自由尊重
- 4.1 健康安全

1. 疏远的关系
2. 人机之融合
3. 绝对的代理
4. 错觉与欺骗


VISITOR
反正别人说话我也不太信
I don't have much faith in words myself.

人际互动



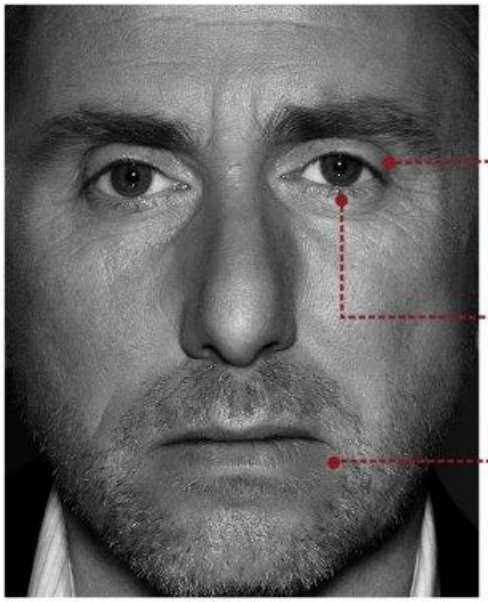
surprise

Lasts for only one second:

① eyebrows raised

② eyes widened

③ mouth open



sadness

① drooping upper eyelids

② losing focus in eyes

③ slight pulling down of lip corners

人际互动

- 人际互动指的是：两个或两个以上个体之间的任何一种**社会遭遇**（social encounter）
- 人际互动可以是正式的也可以是非正式的
- 哪怕是许多漫无目的的互动（如身体语言、面部表情、姿态）都会让人们自得其乐，并完成日常生活各种任务。

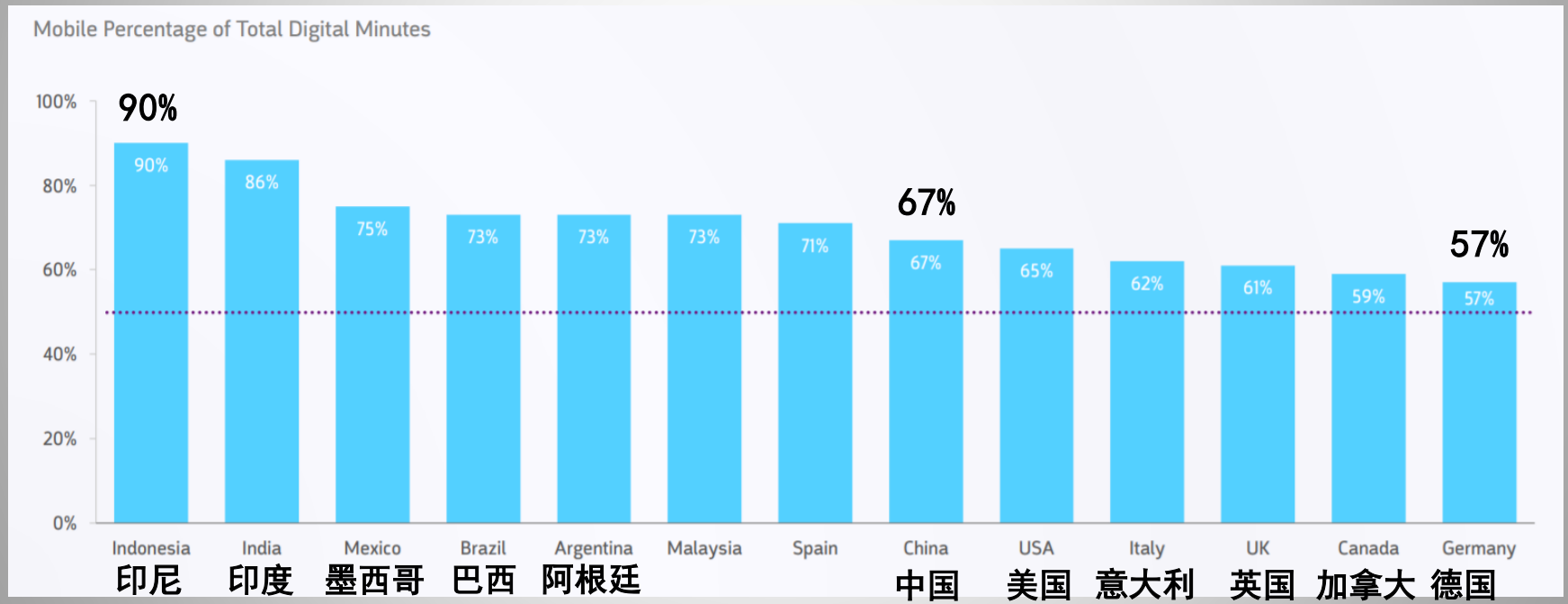
人物 · 观点

Thomas Ploug 【丹麦】
- 信息技术伦理学家，奥尔堡大学教授

> **人们的互动和伦理行为在现实与虚拟空间中存在重要差异**

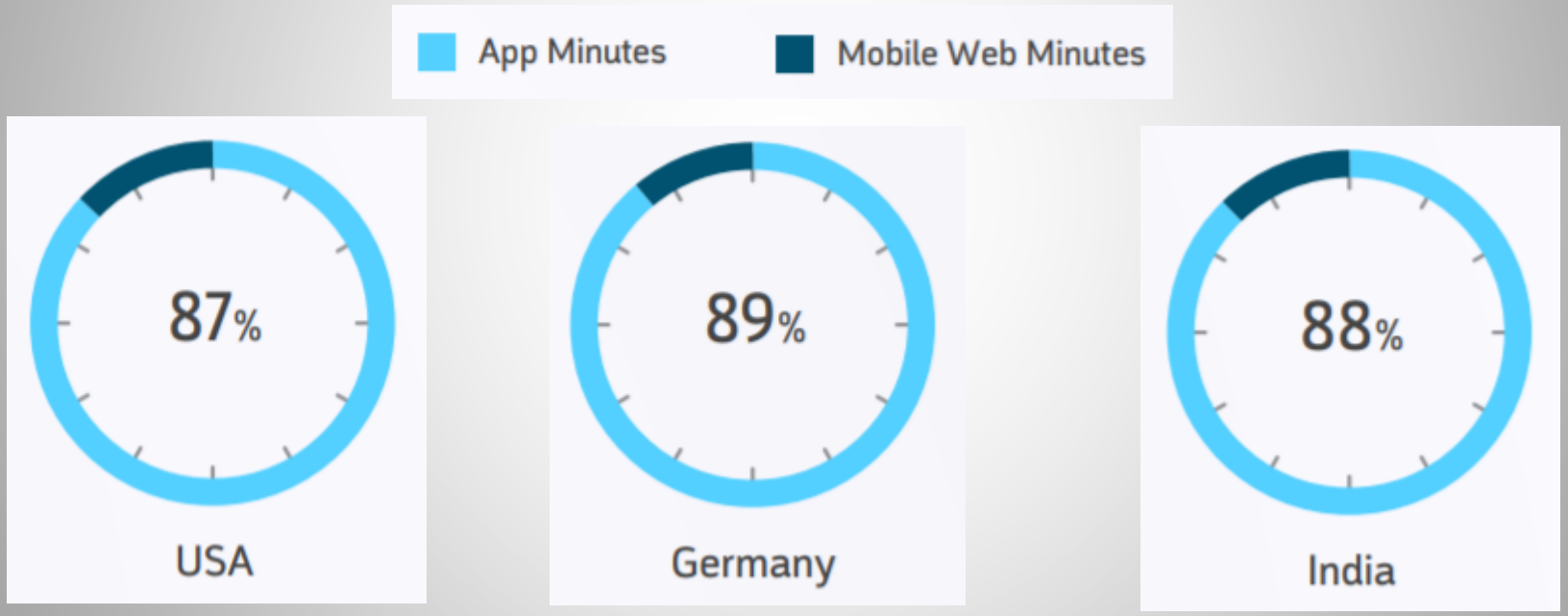
Ploug 认为这种差异会影响人们的道德立场

据调查，在包括中国在内的13国中，移动端占据超过50%的在线时长



Source: comScore Mobile Metrix, May 2017

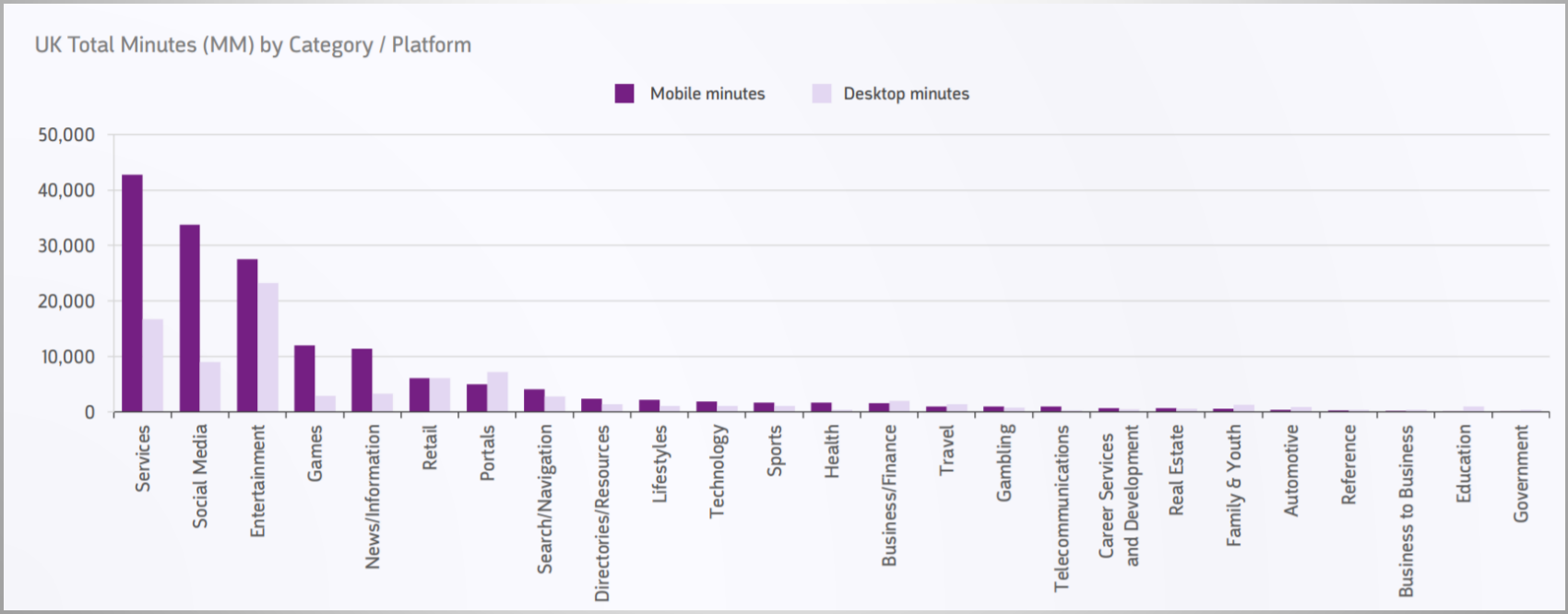
在所调查的13个国家中，80%的时间都在App使用上



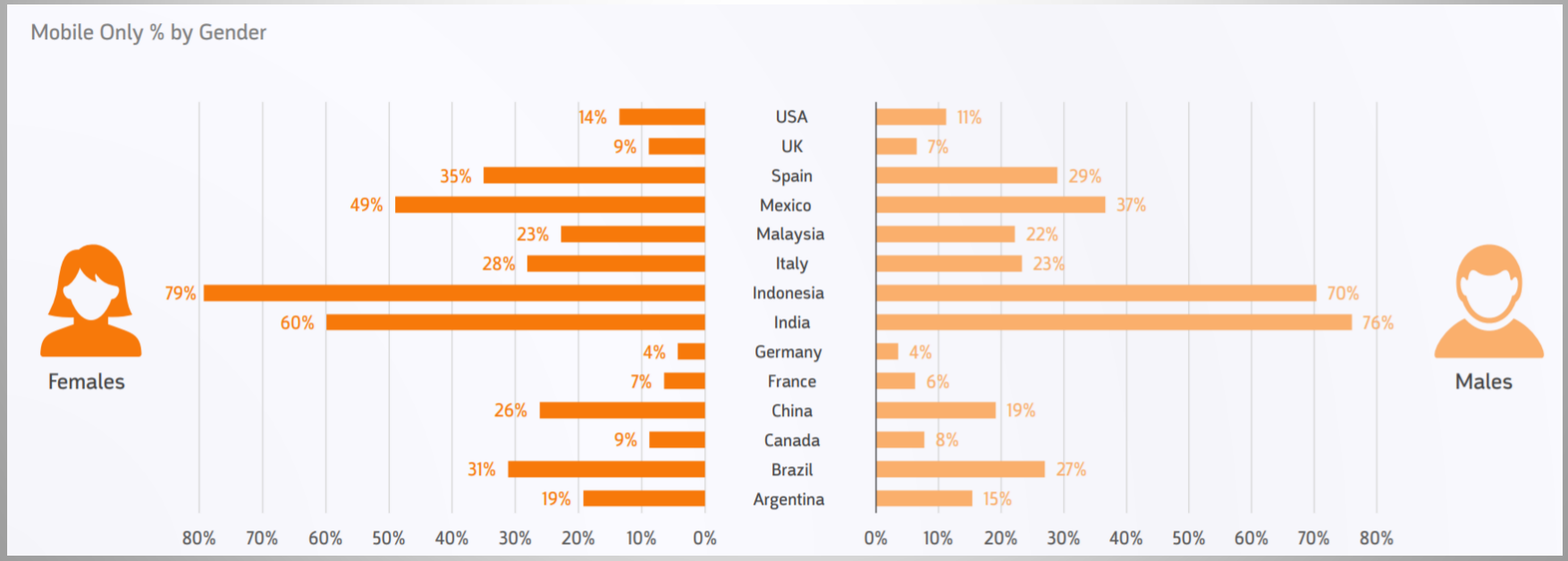
Source: comScore Mobile Metrix, May 2017

More than ¼ of global users are now mobile only

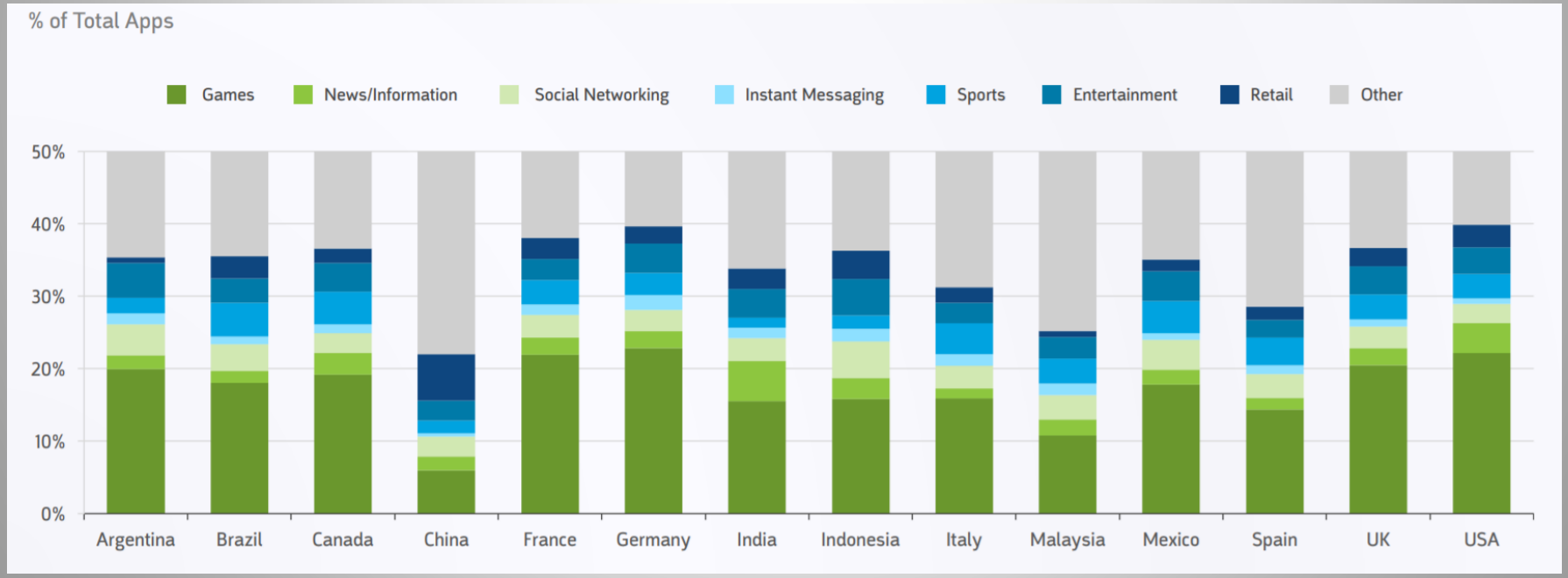
- Overall mobile share is driven by highly mobile categories
- Services (including instant messaging), entertainment and social drive up overall share of mobile minutes



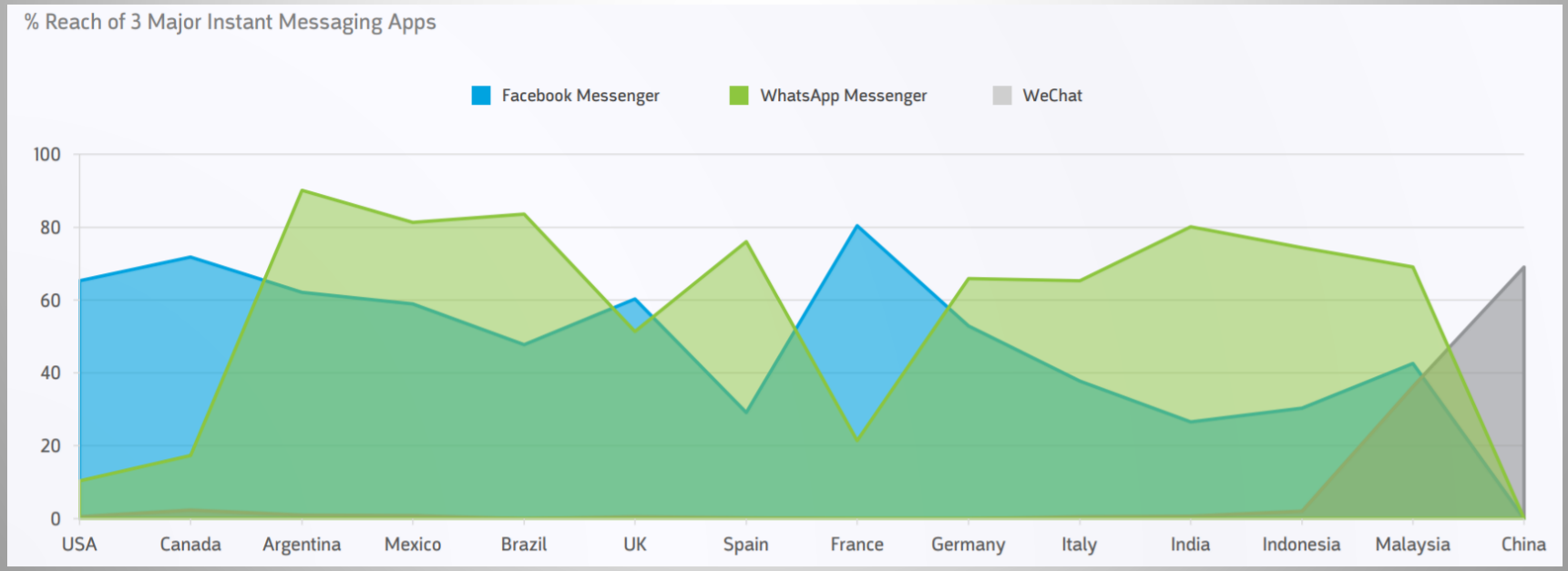
- Females are more likely to be mobile only internet users
- India was the only market where male users were more likely to be mobile only



- The most significant number of apps are games…
- In almost every market, the category with the largest number of apps is gaming



- Adoption almost binary for Instant Messaging Apps
- Despite a relatively small number of apps, markets generally have one established category leader



计算机技术对互动的影响

- 相比现实世界：
  - 网络环境中的各种证据缺乏可信度
  - 网络世界中的争论和不满更高更多
  - 人们表达异见的方式也更激烈激进

有必要弄清楚，为什么网络环境会导致不同的伦理标准，中间的过程机制是怎样的，未来网络互动会有怎样后果

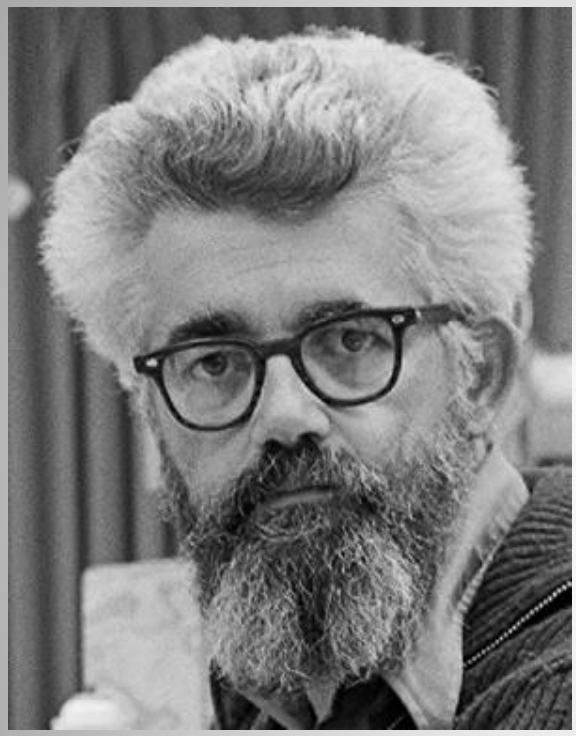
Course Outline

案例总览

- 4.6 产权利益
- 4.5 真实可信
- 4.4 公平公正
- 4.3 隐私保护
- 4.2 自由尊重
- 4.1 健康安全

1. 疏远的关系: 异化的加深
2. 人机之融合
3. 绝对的代理
4. 错觉与欺骗

人物 · 观点



John McCarthy【美】(1927-2011)
- 计算机科学家与认知学家，图灵奖得主

> I don't see that human intelligence is something that humans can never understand.

1955年John McCarthy率先提出了术语人工智能 “Artificial Intelligence”

图灵测试

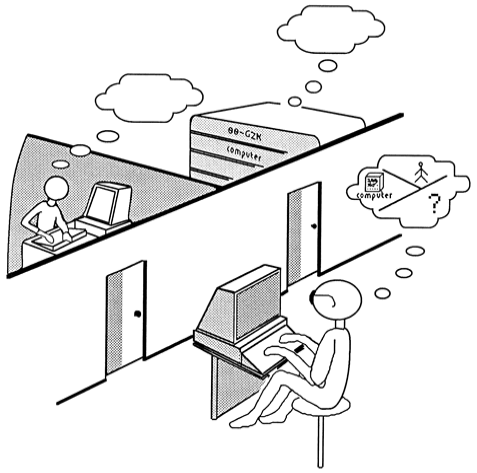
The Turing Test (Alan Turing, 1950)

I.—COMPUTING MACHINERY AND INTELLIGENCE

By A. M. Turing

1. *The Imitation Game.*

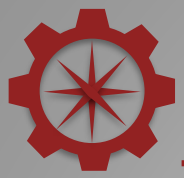
I propose to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, 'Can machines think?' is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed



人类裁判通过随意提问判断被测试者是人是机。若人类判断错误，则机胜利。如果能超过30%糊弄人类，则可以被认为有智能。

与AI对话

2014年6月8日，计算机尤金·古斯特曼成功让人类相信它是一个13岁的男孩，成为有史以来首台通过图灵测试的计算机



与AI对话

Artificial intelligence / Machine learning

OpenAI's new language generator GPT-3 is shockingly good—and completely mindless

The AI is the largest language model ever created and can generate amazing human-like text on demand but won't bring us closer to true intelligence.

by **Will Douglas Heaven** July 20, 2020

人物 · 观点

John Searle【美】(1932-)
- 哲学家，语言学家，加州大学教授

> Strong AI has little to tell us about thinking, since it is not about machines but about programs

哲学家希尔勒认为“强人工智能”需要复制人类大脑的内部因果关系

中文房子问题

The Chinese room argument (John Searle, 1980)

房间里的人可以让任何房间外的人以为他会说流利的中文。塞尔论证：“语法不等同于语义”及“复制不等同于模拟”

人机智能

三元空间群智认知离不开人机协同互补，是一个反馈迭代的过程

认知是人类的推理，计算是机器的推理

验证码中的人机智能

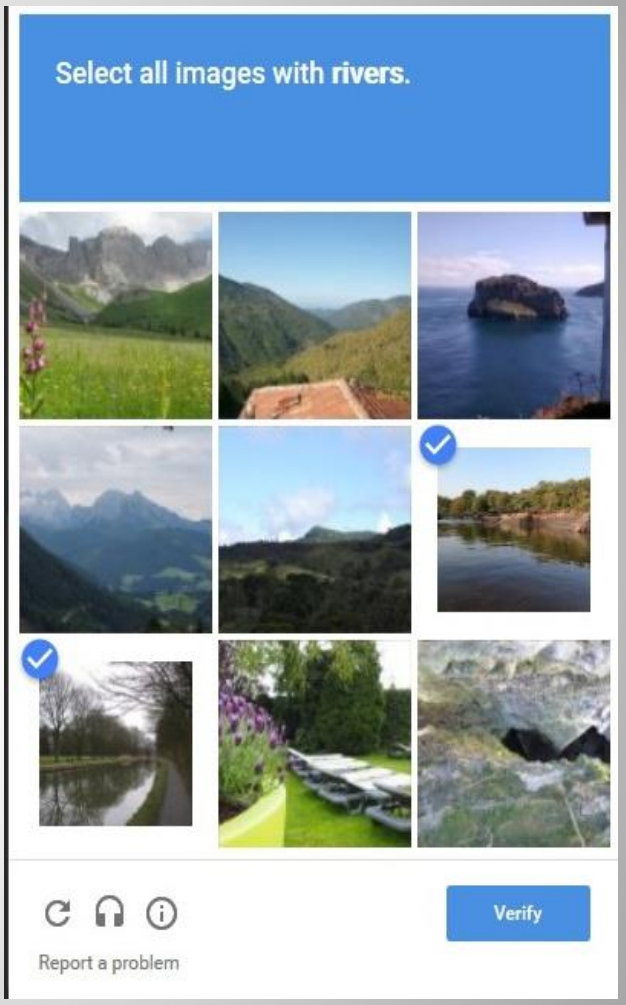
reCAPTCHA: Human-Based Character Recognition via Web Security Measures

Luis von Ahn,* Benjamin Maurer, Colin McMillen, David Abraham, Manuel Blum

CAPTCHAs (Completely Automated Public Turing test to tell Computers and Humans Apart) are widespread security measures on the World Wide Web that prevent automated programs from abusing online services. They do so by asking humans to perform a task that computers cannot yet perform, such as deciphering distorted characters. Our research explored whether such human effort can be channeled into a useful purpose: helping to digitize old printed material by asking users to decipher scanned words from books that computerized optical character recognition failed to recognize. We showed that this method can transcribe text with a word accuracy exceeding 99%, matching the guarantee of professional human transcribers. Our apparatus is deployed in more than 40,000 Web sites and has transcribed over 440 million words.

“Solving Problems with Millions of Humans and Computers”

ACM FCRC 2011



让计算融入生活

Ubiquitous and Pervasive Computing

将计算能力嵌入在日常生活的各类物品中，使之具有通讯和计算能力

- Ubiquitous Computing自80年代起成为一个学科
- IBM 自90年底起设立 Pervasive Computing 部门
- Wearable Computing 如今是其下一个子研究方向

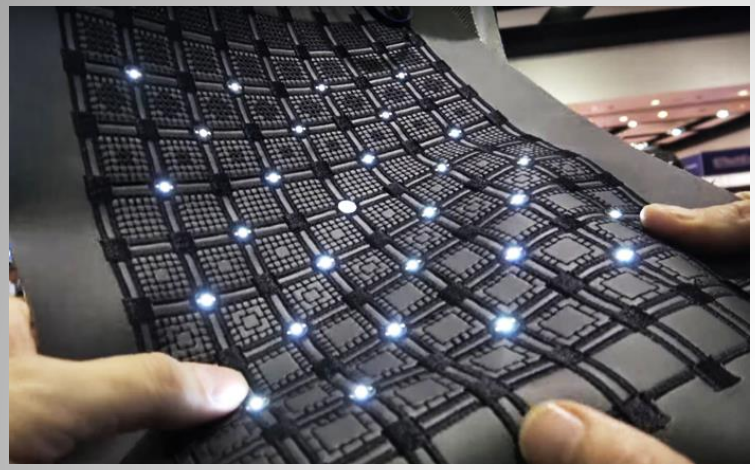
人物 · 观点

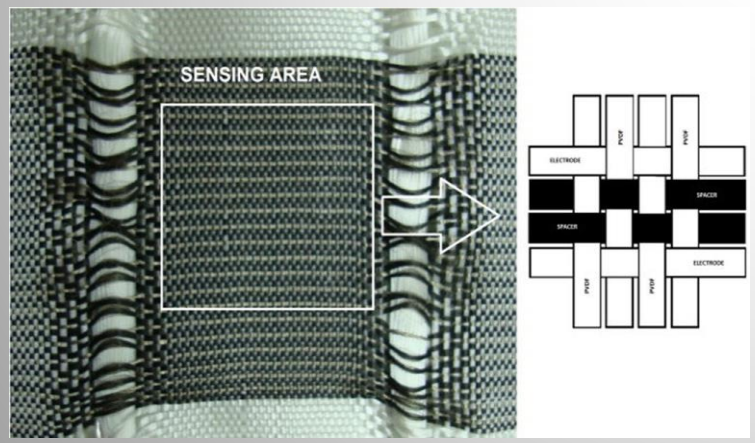
Mark Weiser【美】 (1952-1999)
- 计算机科学家，普适计算奠基人

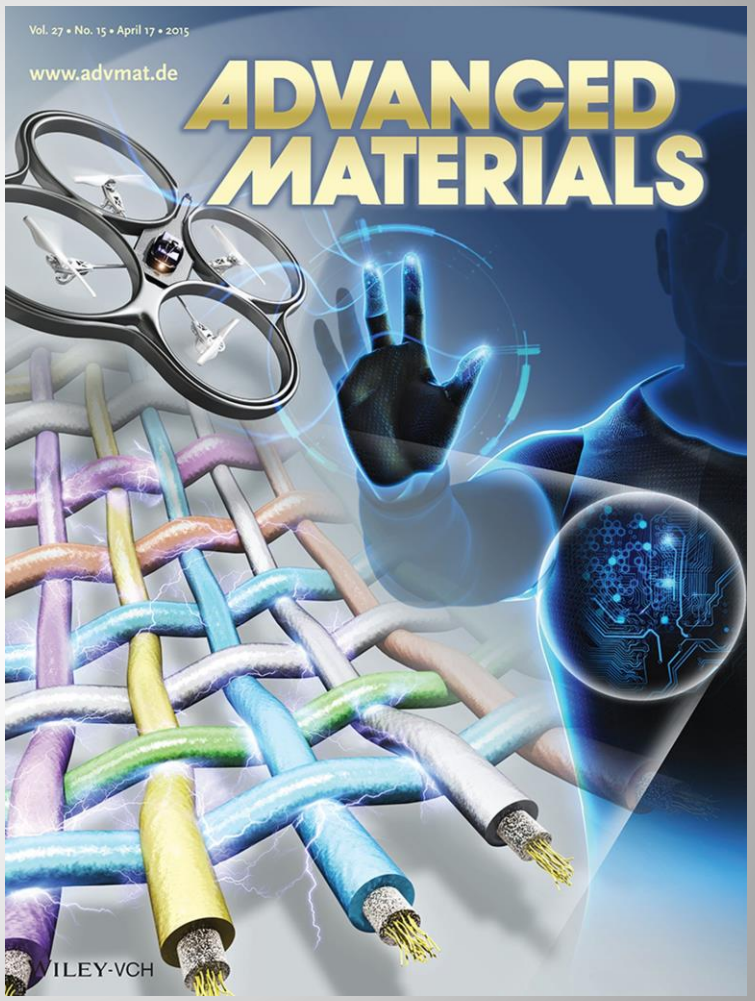
> The most profound technologies are those that disappear

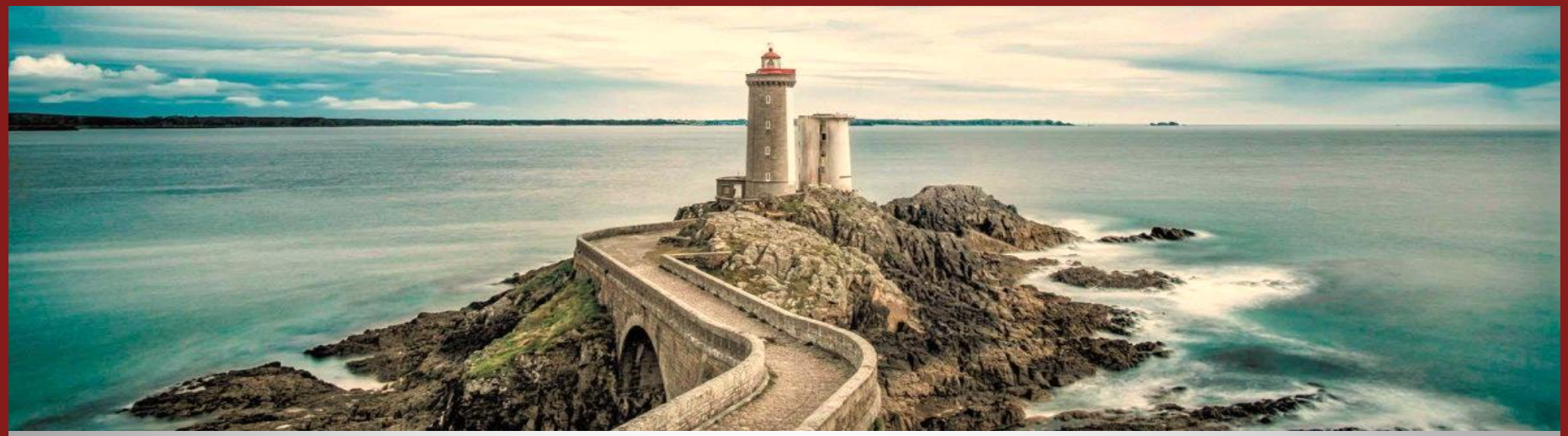
Weiser 认为成功的技术应当是编织入人们的生活中，以至无法分辨

嵌入传感器的功能纤维




SENSING AREA
ELECTRODE
PVDF
PVDF
SPACER
SPACER
ELECTRODE
PVDF
PVDF


Vol. 27 • No. 15 • April 17 • 2015
www.advmat.de
ADVANCED MATERIALS
WILEY-VCH



Course Outline

案例总览

- 4.6 产权利益
- 4.5 真实可信
- 4.4 公平公正
- 4.3 隐私保护
- 4.2 自由尊重
- 4.1 健康安全

1. 疏远的关系: 异化的加深
2. 人机之融合: 有限度参与
3. 绝对的代理
4. 错觉与欺骗

导言

自主计算 Autonomic Computing

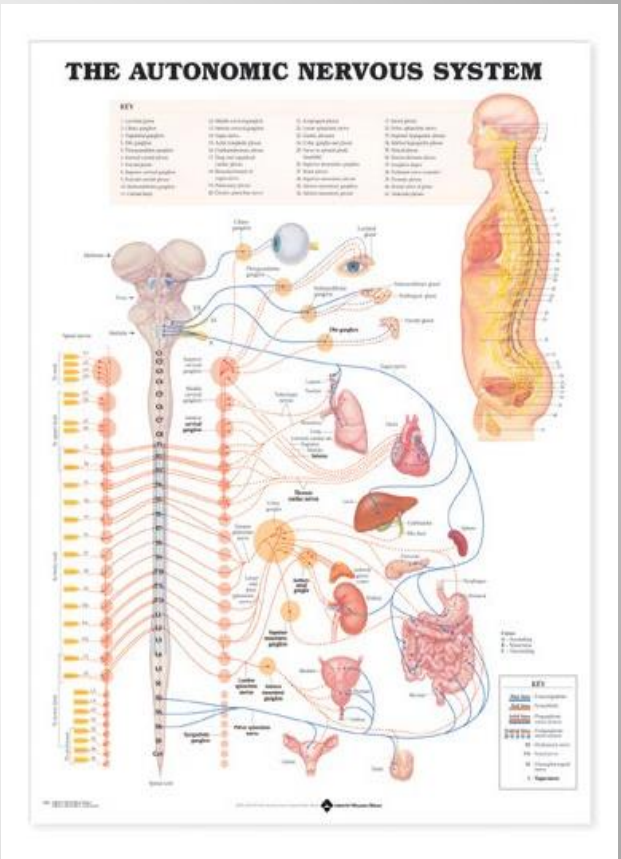
Autonomic Computing is a comprehensive approach to build more automated IT infrastructures that require minimal intervention

AC是IBM公司2001年开始引入的概念

自主性

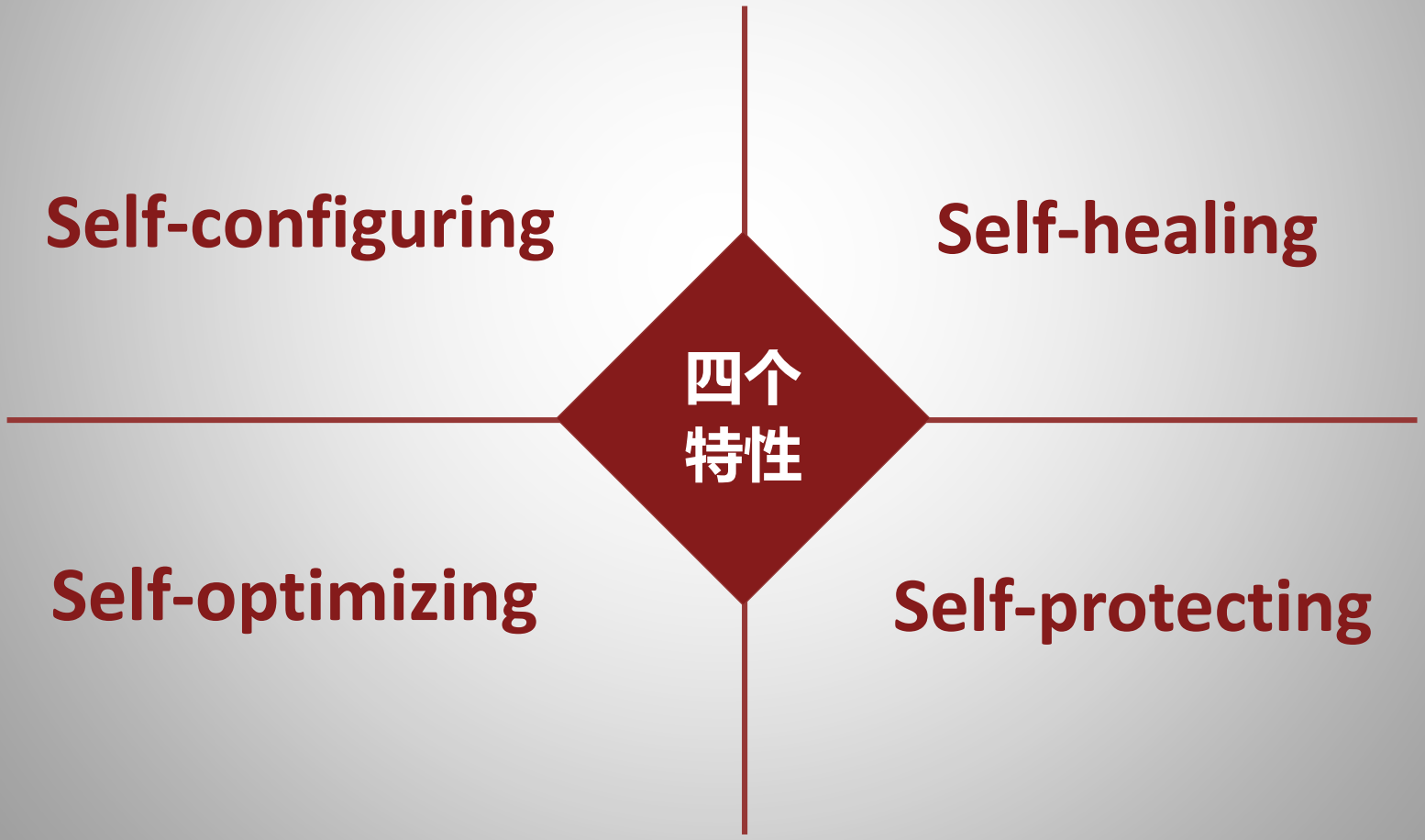
自主计算的自主一词“autonomic”体现仿生性

- 自主神经系统：Autonomic Nervous System
  - 许多人体功能不需要我们主动去管理，比如瞳孔大小，呼吸节奏，伤口重建
  - 若没有自主的管理，人们可能就疲于管理自身对环境的适应而无法生活创造了。



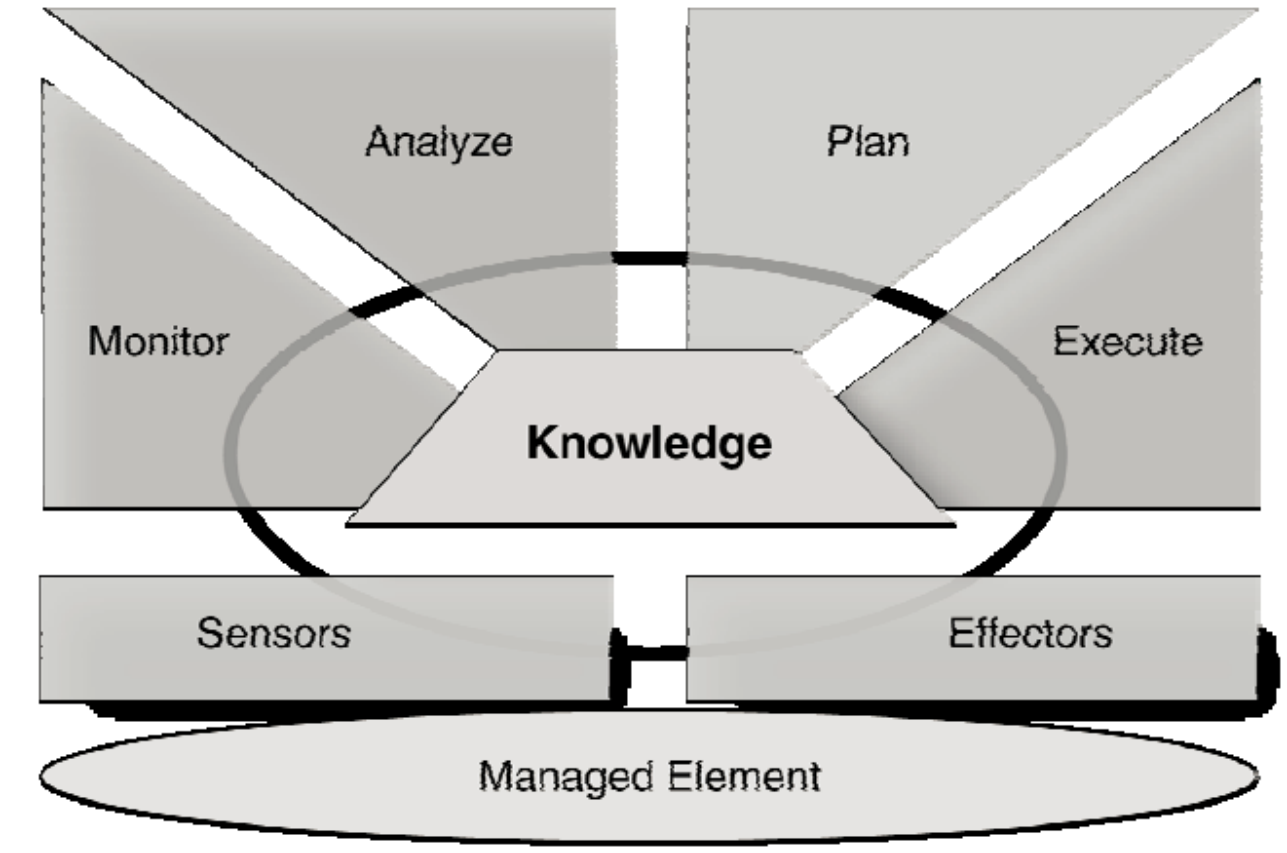
自主计算系统的功能属性

核心思想是 Self - *

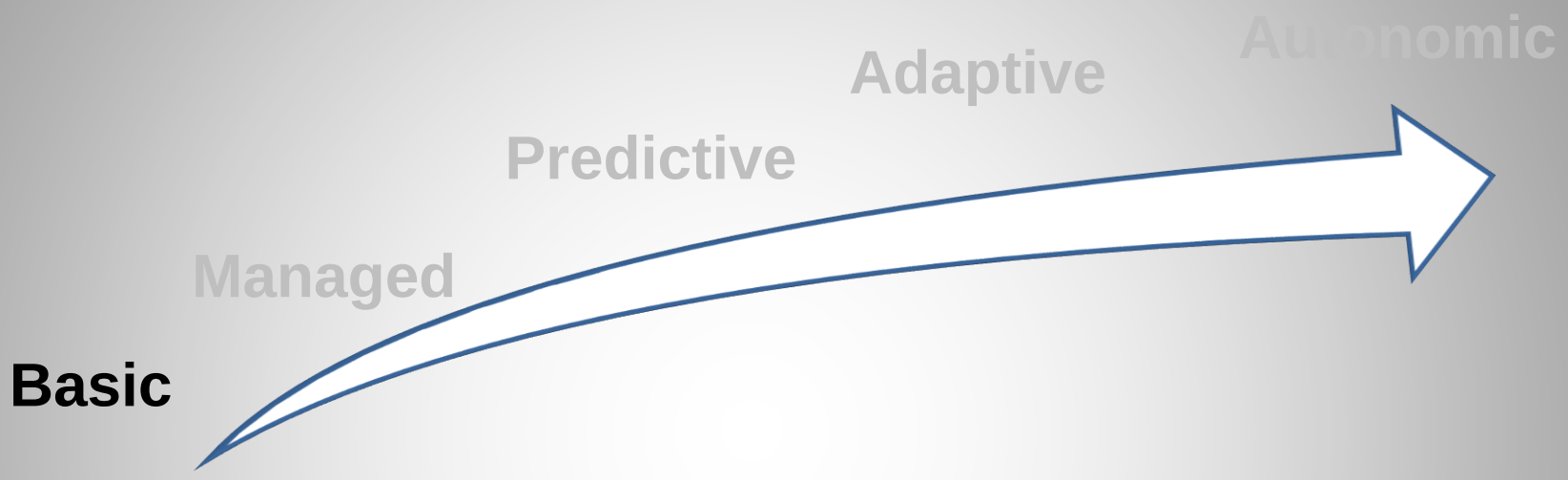


自主计算系统概念模型

An autonomic element manages its own internal state and its interactions with its environment (i.e., other autonomic elements)

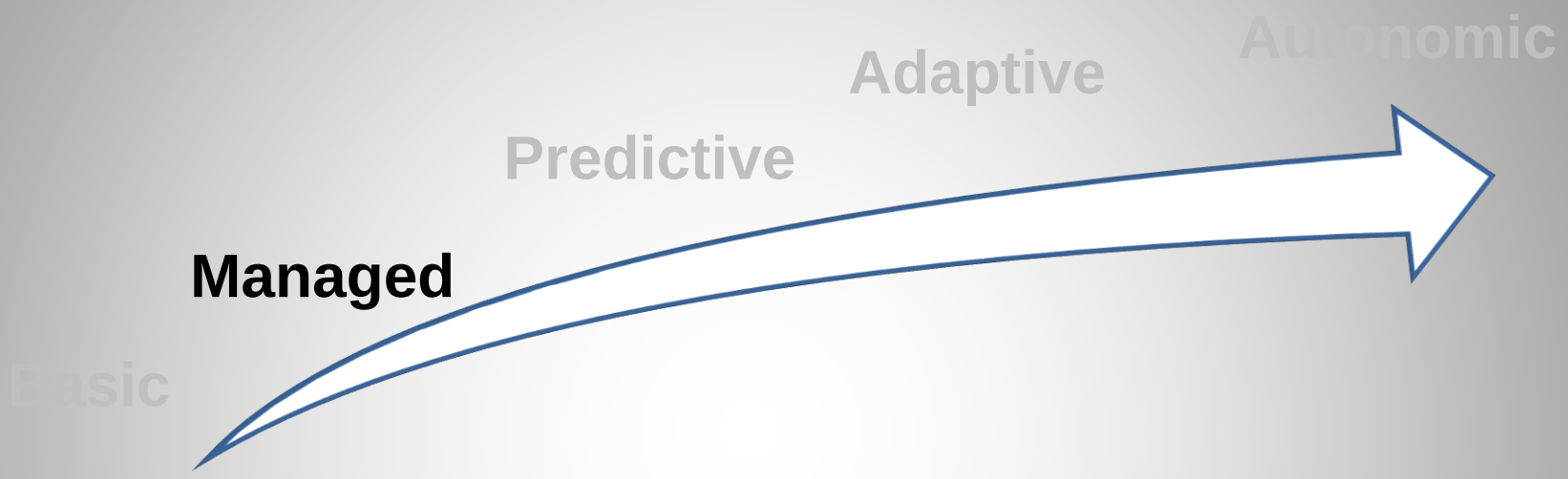


自主性的演变路径



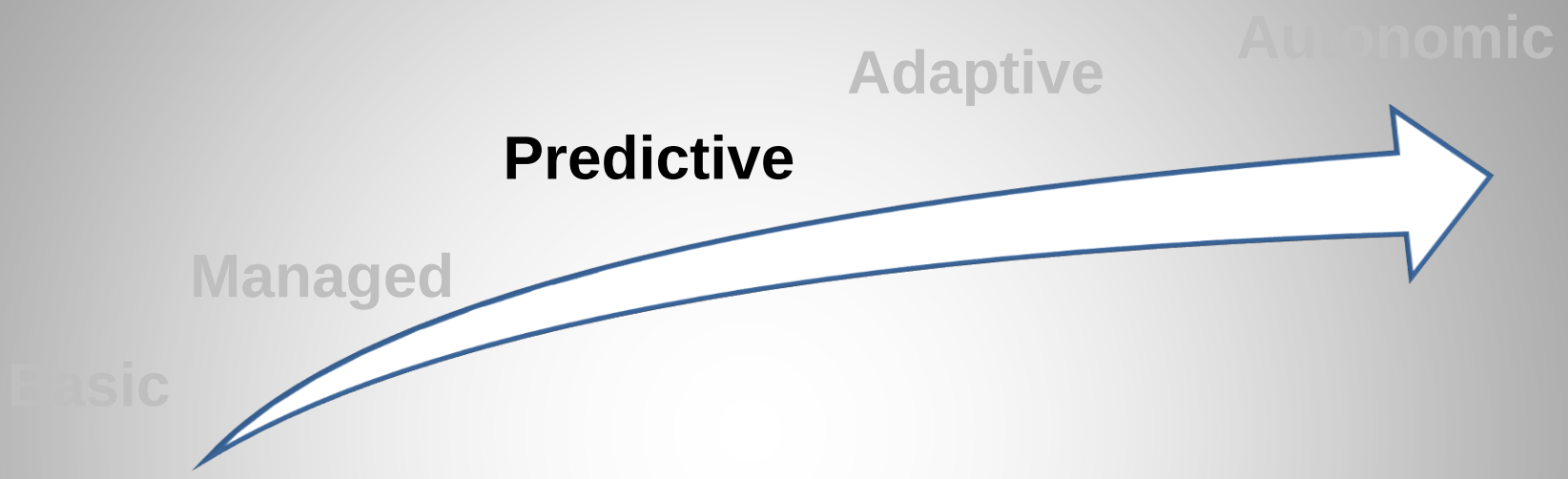
The basic level represents the starting point where a significant number of IT systems are today. Each element of the system is managed independently by systems administrators who set it up, monitor it, and enhance it as needed.

自主性的演变路径



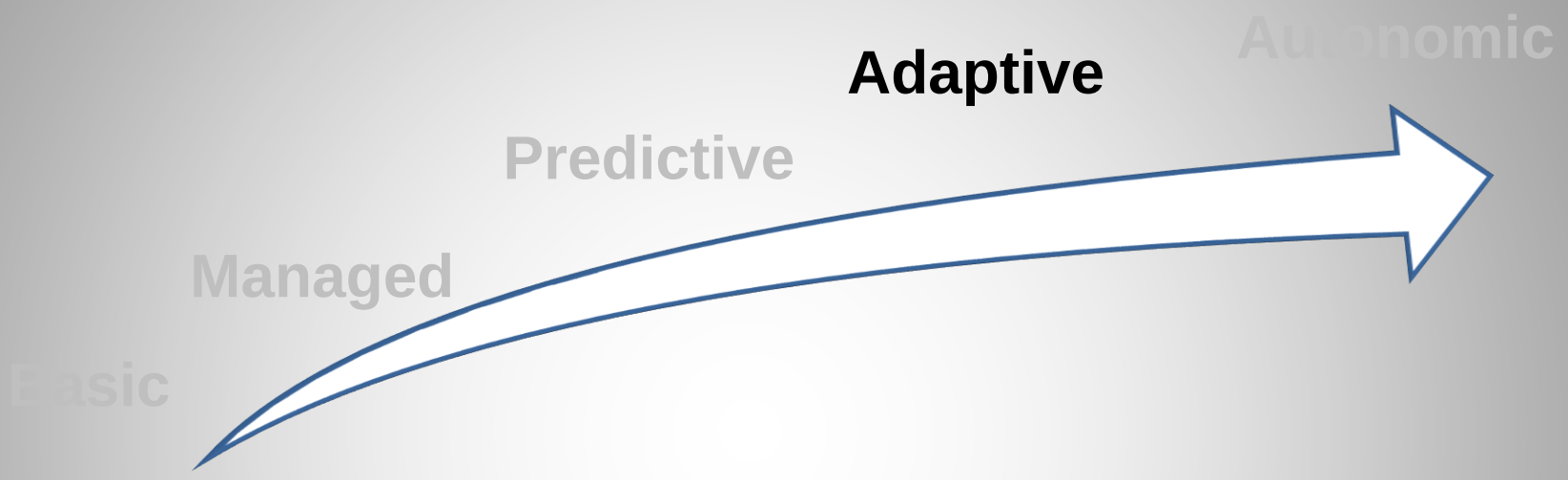
At the managed level, systems management technologies are used to collect information from disparate systems into one, consolidated view, reducing the time it takes for the administrator to collect and synthesize information.

自主性的演变路径



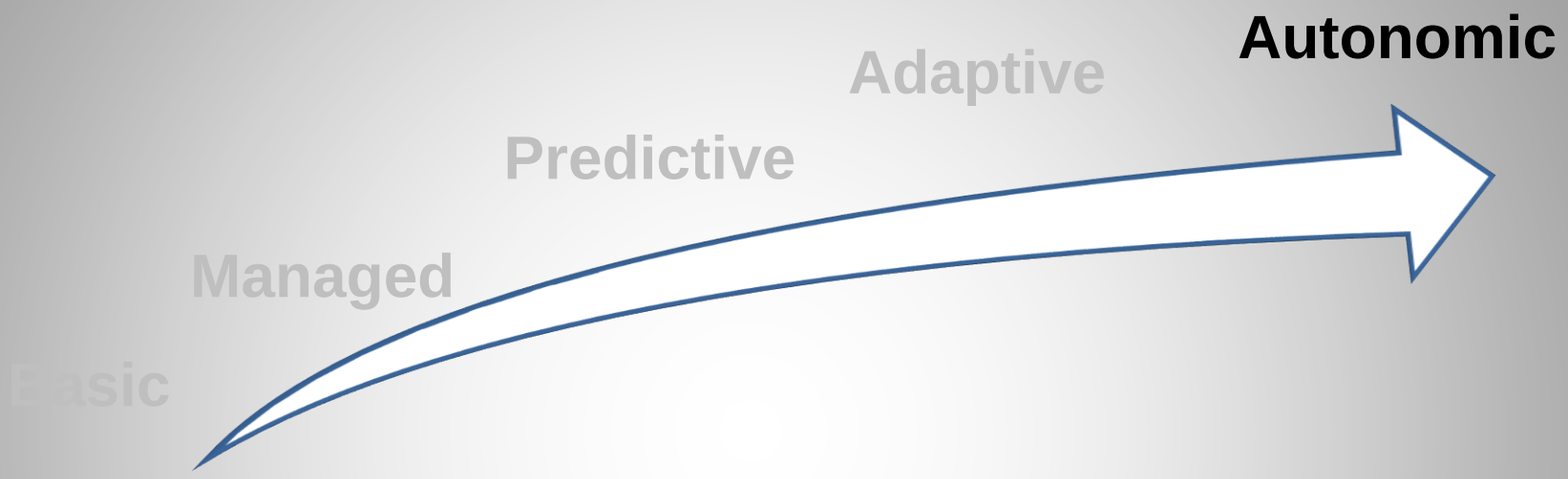
At the predictive level, new technologies are introduced. The system itself can begin to recognize patterns, predict the optimal configuration and provide advice on what course of action the administrator should take.

自主性的演变路径



The adaptive level is reached when systems can not only provide advice on actions, but can automatically take the right actions based on the information that is available to them on what is happening in the system.

自主性的演变路径



Finally, the full autonomic level would be attained when the system operation is governed by business policies and objectives. Users interact with the system to monitor the business processes, and/or alter the objectives.

Autonomic Computing: A Threat?

- Once systems become autonomous, by definition they will have a "**mind of their own**."
- Autonomic computing may lead to true complex adaptive systems whose ultimate behavior is very **difficult to predict** from initial conditions
- When allowing systems to autonomously operate, we must be sure we are able to **bound the consequences** and that we are willing to pay the potential price.

对伦理学的挑战

- 通过普适系统来开发我们的日常生活，将会对伦理学探讨的三个条件产生巨大影响
  - 人在其中进行活动的“**现实世界的可决定性**”
  - 承担责任过程中的“**行为主体的同一性**”
  - 以及使负责人行为成为可能的“**遴选抉择**”

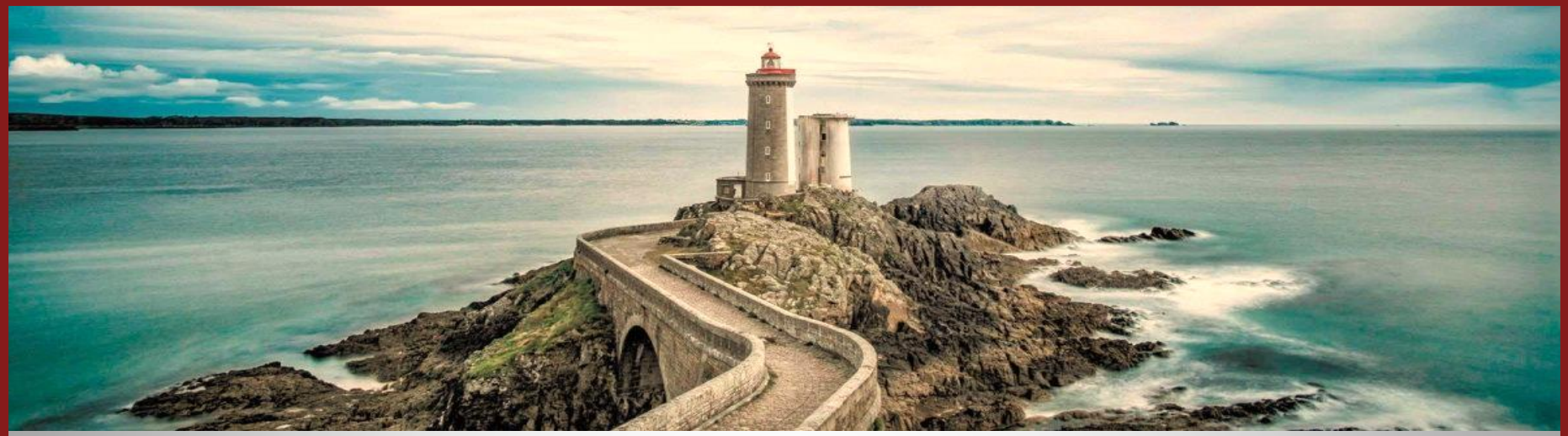
对伦理学的挑战

- 实践证明，作为现实世界基本标志的抗拒能力的丧失，乃是一个根本性的问题。

- 一个不让人了解接口与抗拒能力的技术既无法受到检查也无法收到控制。

对伦理学的挑战

- 在使用普适系统时，是否能够避免家长制效应和剥夺别人行为能力，这个问题很重要。
- 德国科学基金会一个研究项目Nexus，旨在缓和诸如剥夺用户行为能力和技术的不可控性的一种尝试。
- 应该根据合理性原则给予用户对系统**控制的可能性**，以及对系统设置进行**干预的可能性**



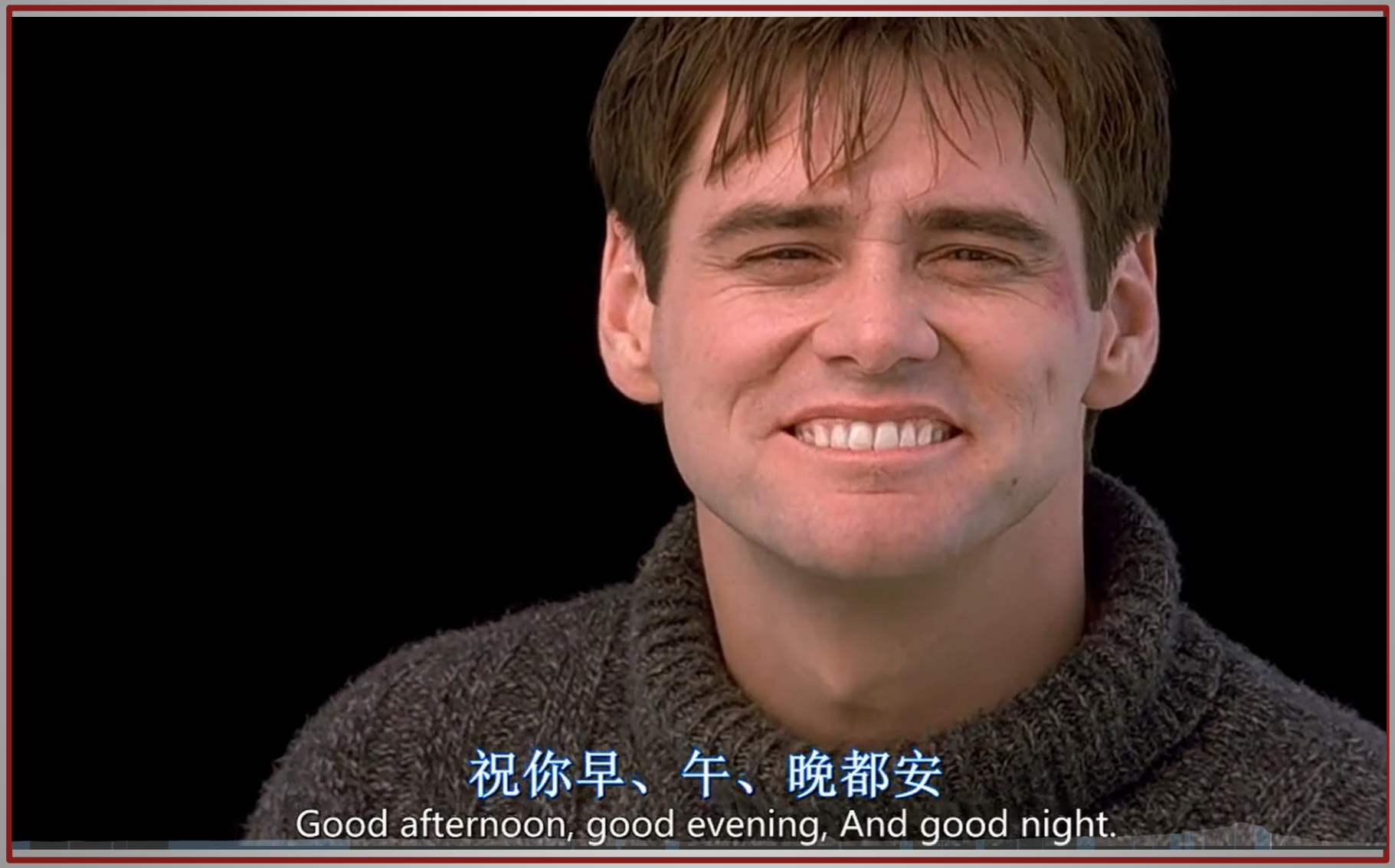
Course Outline

案例总览

- 4.6 产权利益
- 4.5 真实可信
- 4.4 公平公正
- 4.3 隐私保护
- 4.2 自由尊重
- 4.1 健康安全

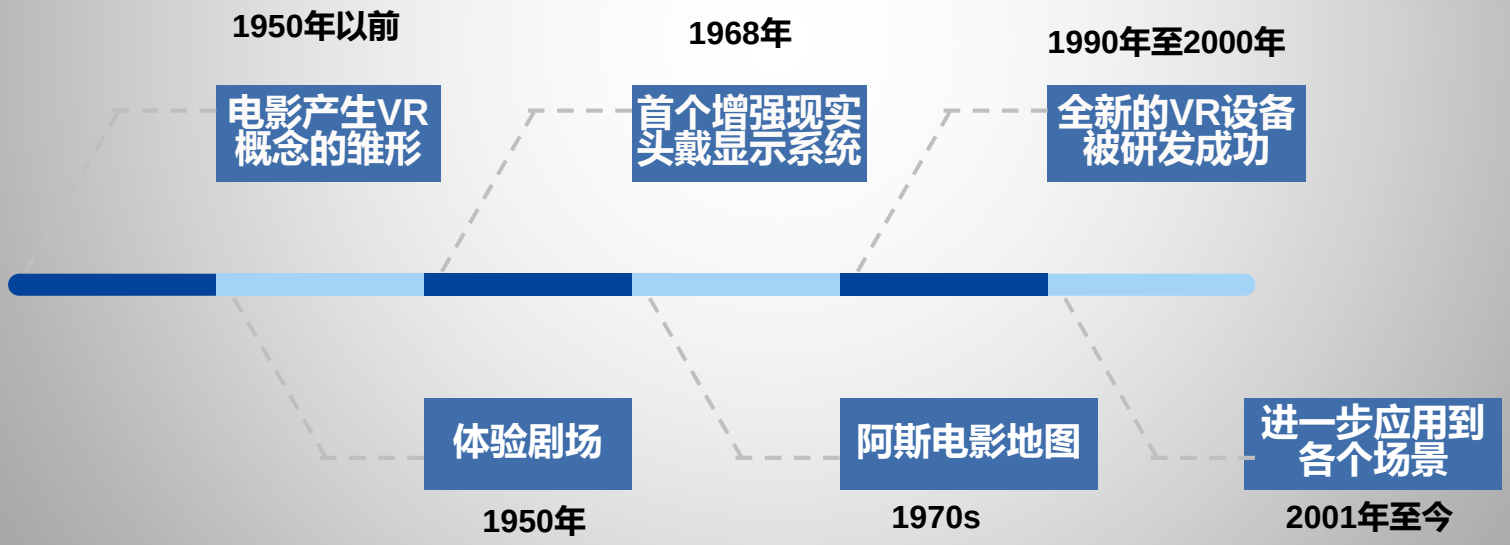
1. 疏远的关系: 异化的加深
2. 人机之融合: 有限度参与
3. 绝对的代理: 可控性问题
4. 错觉与欺骗

导言



虚拟现实：发展历史

VR技术萌生于人们对于虚拟世界的沉浸式体验的需求
VR技术受到广泛关注，获取到软件和硬件方面的支持
VR技术广泛应用到生产生活的各个领域，带来新体验



早期虚拟现实设备

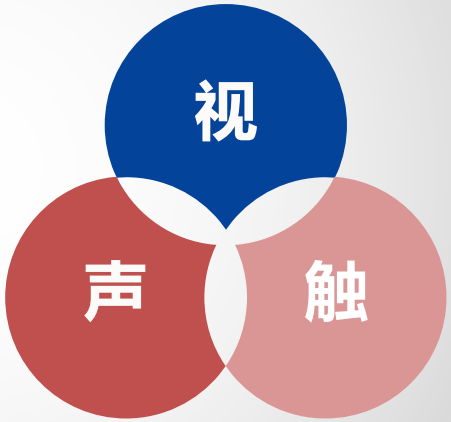
- In the late 1960s Sutherland worked to develop a head-mounted display (HMD)
- it could provide a three dimensional, computer-generated world which the user could view in a normal manner.

虚拟现实：核心技术

- **具体来讲，VR涉及到的核心技术包括：**
  - 计算机图形学
  - 仿真技术
  - 三维建模技术
  - 三维现实技术
  - 动作捕捉技术
  - 环境建模技术
  - 立体显示技术
  - 立体声技术
  - 三维虚拟声音
  - 交互技术
  - 传感器技术
  - 真实感觉时技术
  - 碰撞技术
  - 力反馈技术

VR技术的本质是为用户带来沉浸式的体验

视觉是沉浸式体验的第一感官

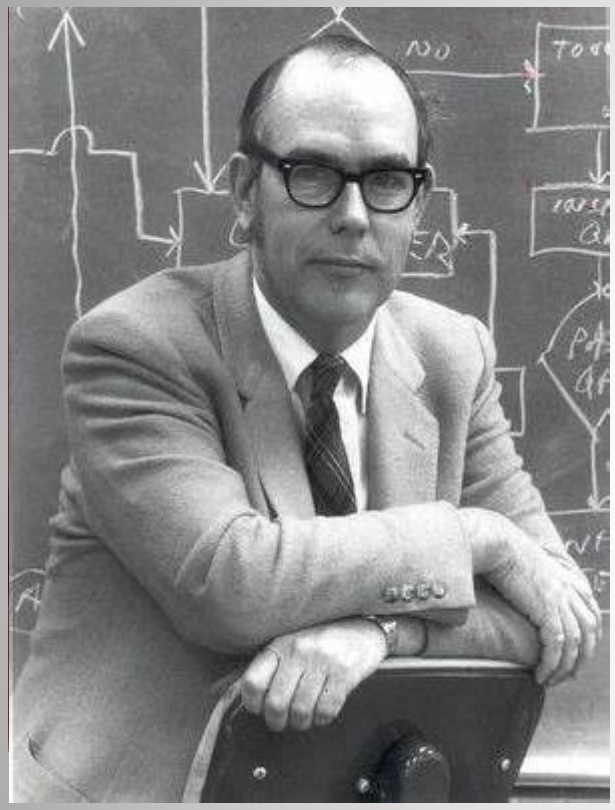


声音增益虚拟世界的立体性

触觉建立反馈系统

VR的核心技术包括了声、视、触三个维度

人物 · 观点



Ivan Sutherland【美】(1938-1895)
- 计算机图形学之父，图灵奖得主

> The screen is a window through which one sees a virtual world

Ivan认为挑战在于让虚拟世界在不同感官维度上都接近真实

讨论

美颜相机、虚拟试衣等应用有何伦理问题?

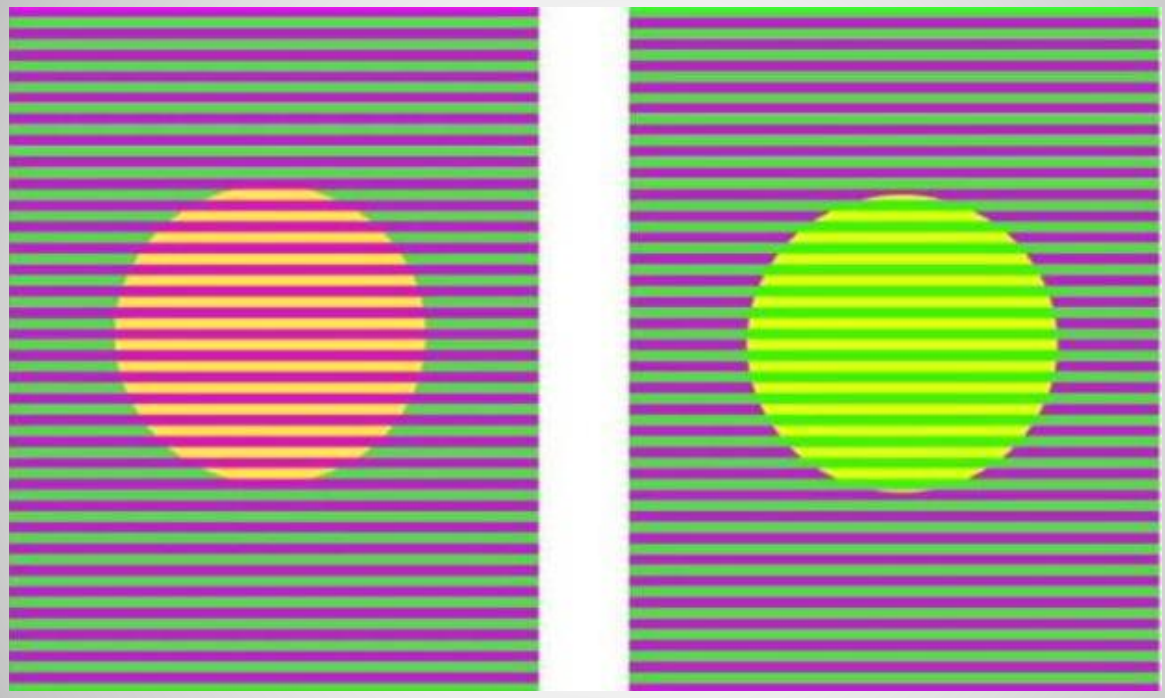
人为什么会说谎?

有记录表明人们一天会说谎1~2次 (Depaulo et al. 2003)

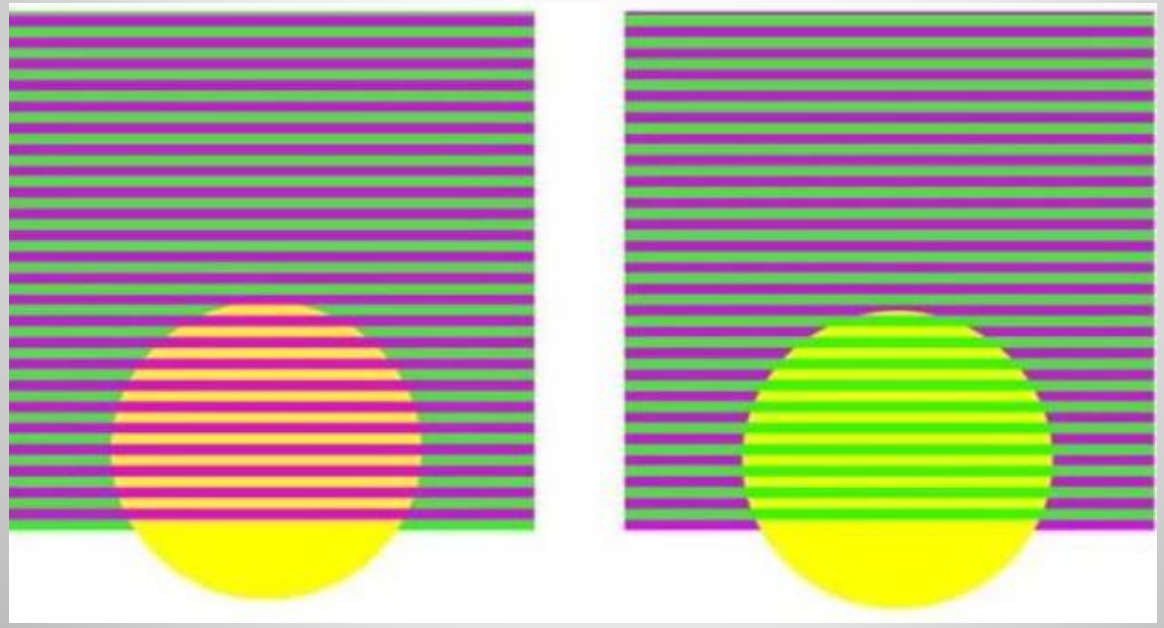
计算机是否应该说谎?

电影星际穿越中，机器人Tars会设置一个幽默指数（置信度）

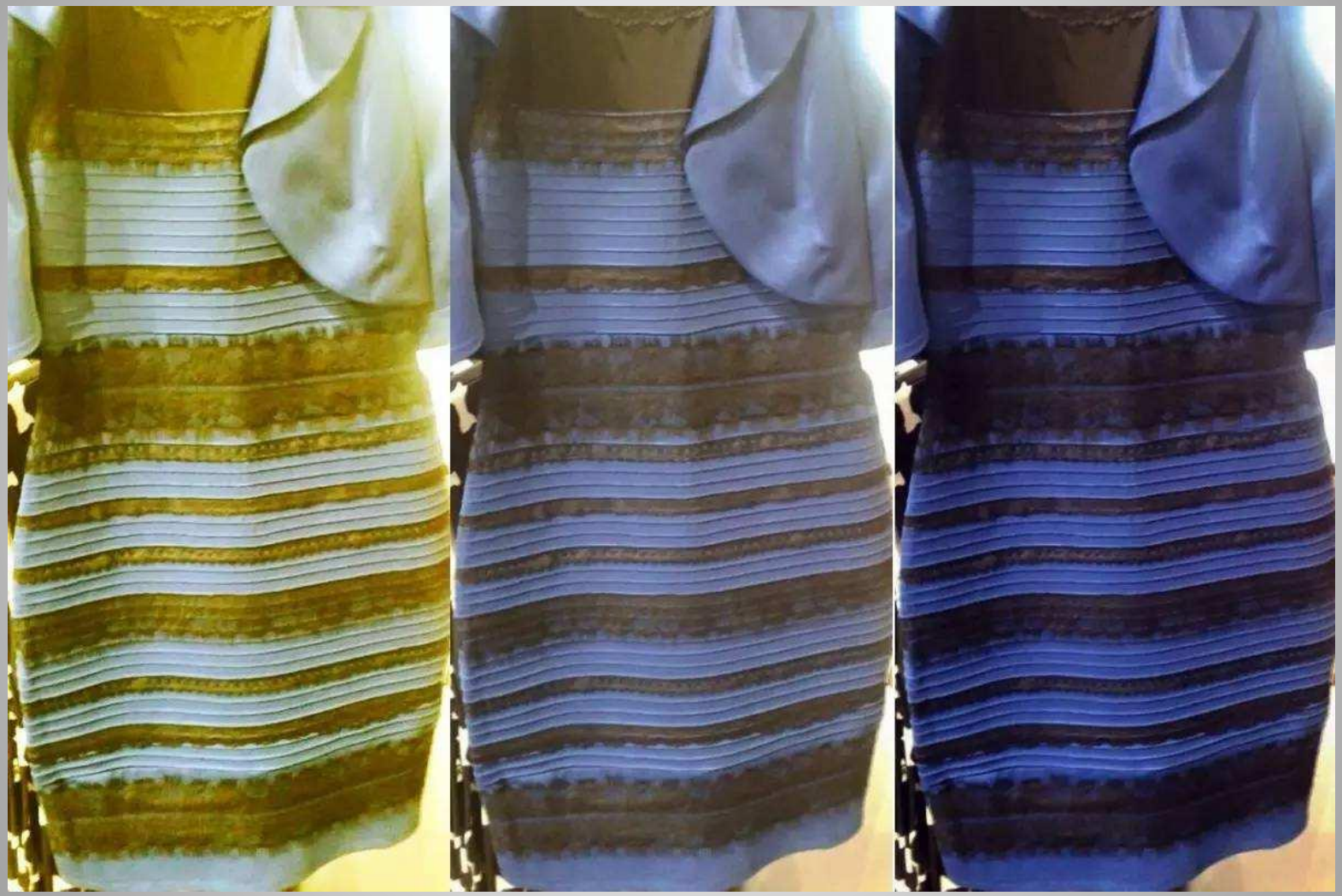
错觉

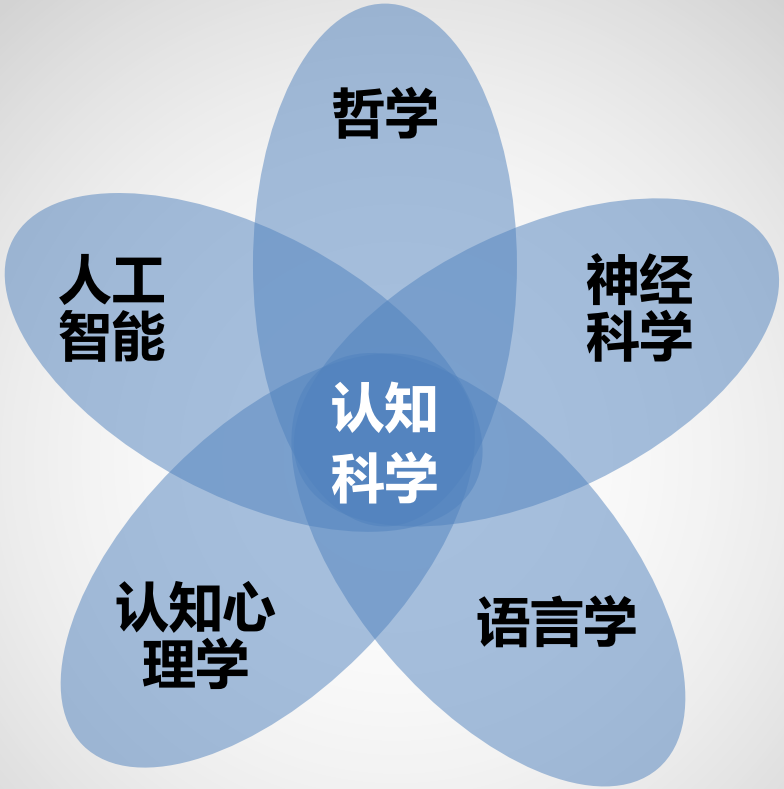


错觉



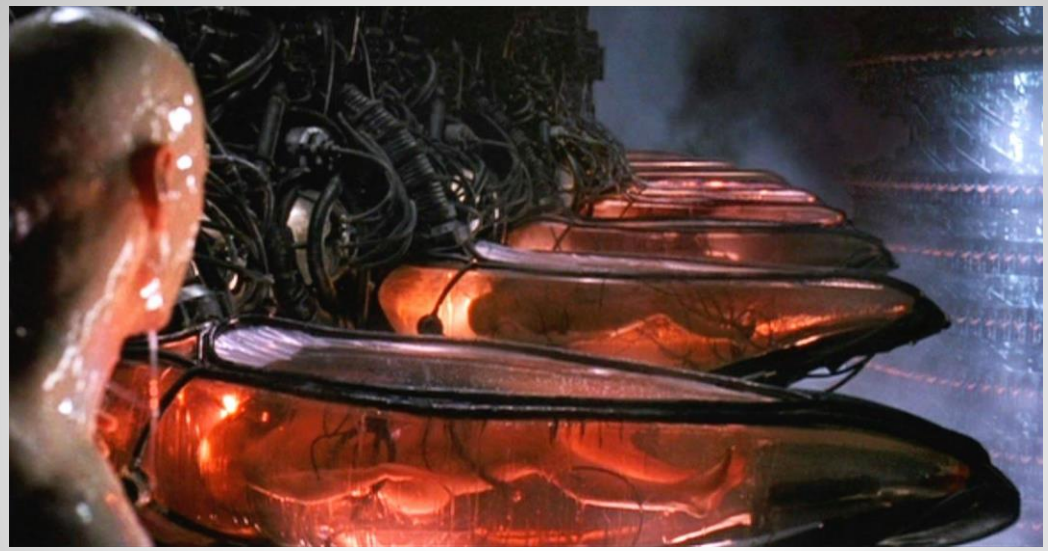
错觉



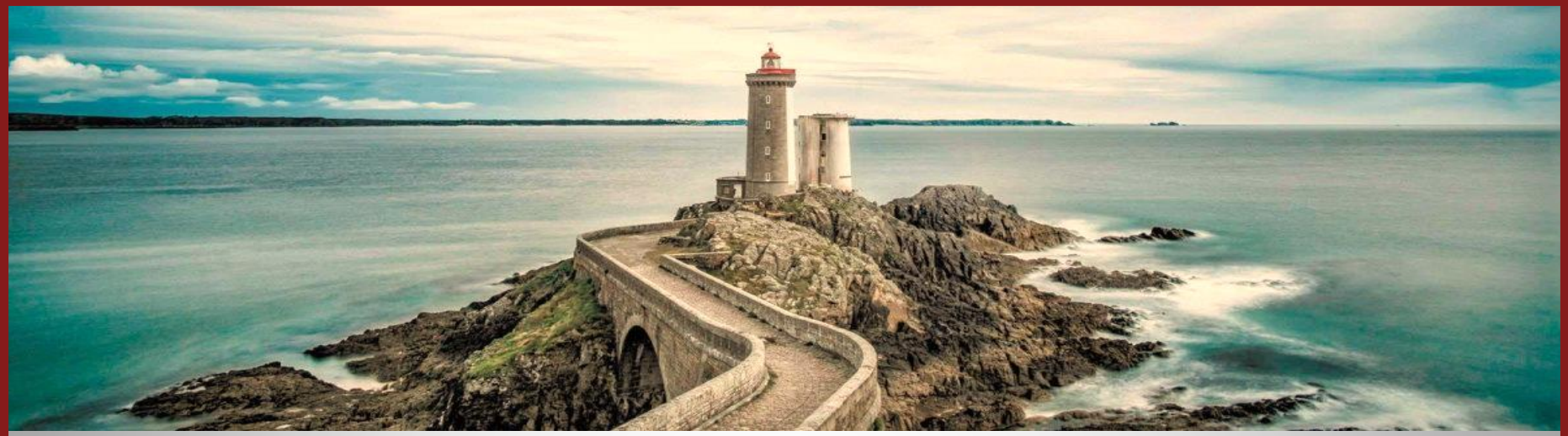


认知科学的领域

黑客帝国的反思



- 黑客帝国中，当人与母体（matrix）联结时，他们或可以选择想要过的生活，但这并不是真实的



Course Outline

案例总览

- 4.6 产权利益
- 4.5 真实可信
- 4.4 公平公正
- 4.3 隐私保护
- 4.2 自由尊重
- 4.1 健康安全

1. 疏远的关系: **异化的加深**
2. 人机之融合: **有限度参与**
3. 绝对的代理: **可控性问题**
4. 错觉与欺骗: **要关注奇点**