

数据科学与工程导论

Introduction to Data Science and Engineering



开放的世界

数据科学与工程职业规划

3 数据隐私与社会问题

数据与人工智能伦理

近40年以来,信息技术社会的演进经历了五大形态

IT信息时代

互联网时代

移 动 互联网时代

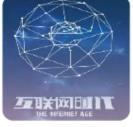
大数据时代

AI时代

IOT时代?

A 互联 B 开放 C 协作 D 演化

Information Technology









· 纽约12 据经 约12 据经决益据 分为已,日数析 报栏: 大代临将于分析

出

• 2010 日村。联以次会全芽 年代它网来技形球 后来是诞第术态的 四浪叠加:

▶大数据

▶人工智能

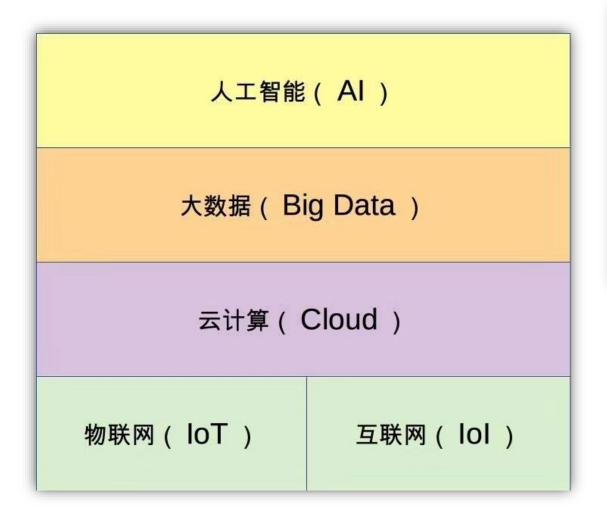
≻物联网

>5G

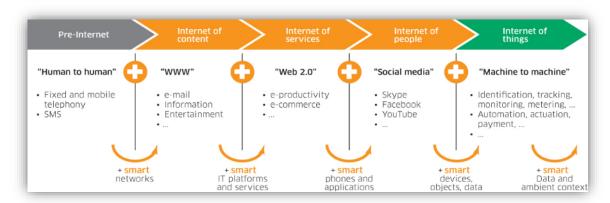
上世纪80年代,PC的快速普及,信息化成为时代主题

· 世代技联迅,信成主纪,术网猛互息为题 90通和技发联共时 • 21十着机信化互成主世年智爆息出联为题的随手,人,+代前随手,人,+代

世界潮流浩浩荡荡

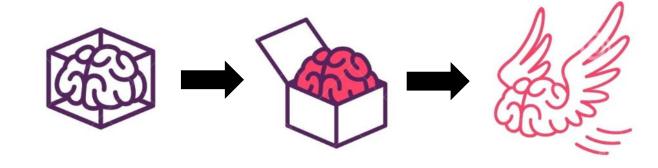








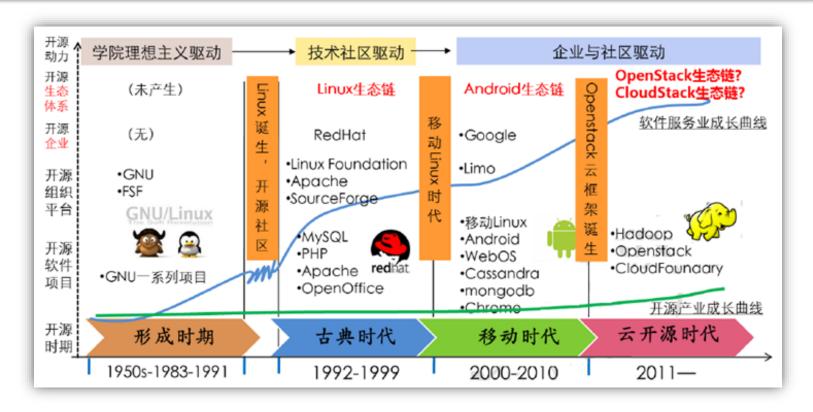
- 新时期的卓越IT人: 应用驱动创新的开放成长体系
 - 开放的思想
 - 开放的问题
 - 开放的教育
 - 开放的平台
 - 开放的技能



开源技术

Living on the cloud, swimming in the (open source) pool

开放



- 1. Git
- 2. Hadoop
- 3. Spark
- 4. ÖpenStack
- 5. KVM
- 6. Xen
- 7. Docker
- 8. Ceph
- 9. Mininet
- 10.OpenDaylight
- 11.Node.js
- 16.Kubernetes
- 17.Mesos
- 18.TensorFlow

- 今天已经不是开源和闭源之间的选择与博弈,而是大面积的开源与开源项目之间的竞争,这里面包含了技术、社区、人才等众多方面的综合因素。所谓,得开放者得天下!
- 例如, Kubernetes社区战略性的全方位吸引开发者的策略令人叹为观止!





开源技术发展历程

- 权威合法组织
- 政府接口中心
- 财/才 资源中心
- 法律,规范中心
- 厂商接口中心
- 委员会形式管理
- 代码审核权控制
- 代码递交权控制
- 成熟的全球化联合开发 协作模式
- 核心开发成员由对口基

FREE SOFTWARE





顶级开源软件基金会







金会聘用

上

社

区

Linux 内核











MySQL







其他 开发社区

开源软件产品化

 以订阅模式交付 商业开源产品

• 与主流社区保持 紧密联系和互动

下 游 生

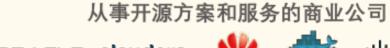
态





















ORACLE cloudera



开源技术生态体系

自然界和人类社会中的典型群体智能现象



互联网的出现为人类群体智能的进一步发展提供崭新的技术支撑

协作

显著 变化

突破时空、超大规模的松散个体可以通过互联网 进行更加方便灵活的显式或隐式交互

- 互联网的不断发展, 促成了互联网群体智能在解决多个领域的问题中展现了神奇的力量:
 - Wikipedia
 - Github
 - Stack Overflow
 - Bitcoin











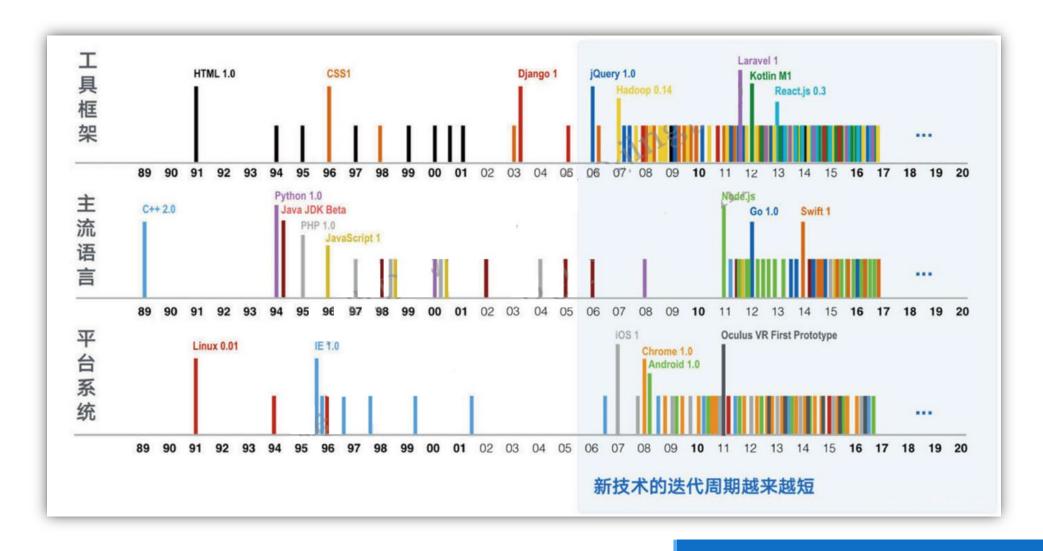
基于协作的互联网群体智能

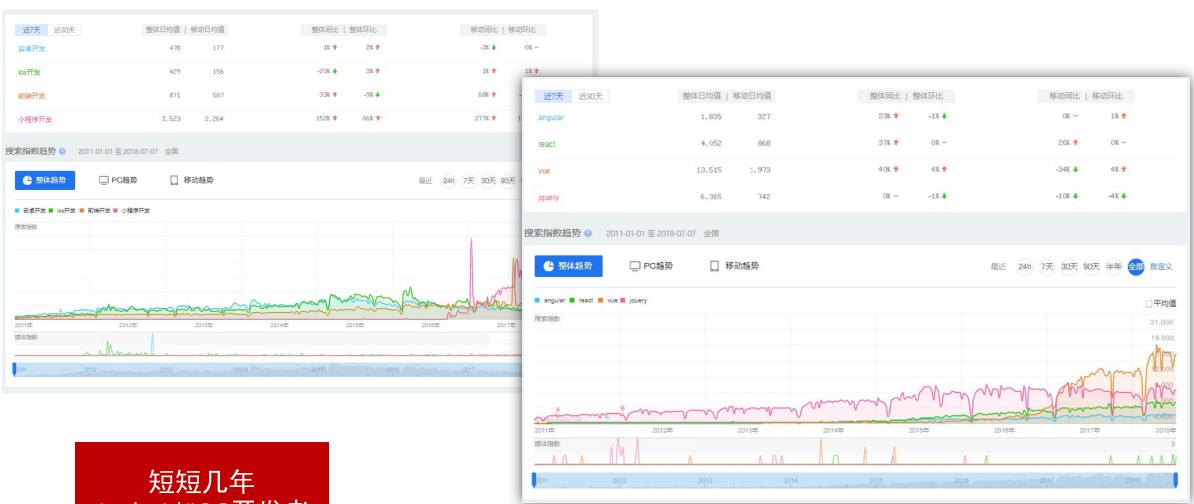
面向协作、提供良好的分享和协作体验、极大降低了参与协作的门槛

- 1. 标准的沟通方式
- 2. 标准的代码管理方式
- 3. 标准的 Web 社交方式
- 4. 开源现在看起来很酷



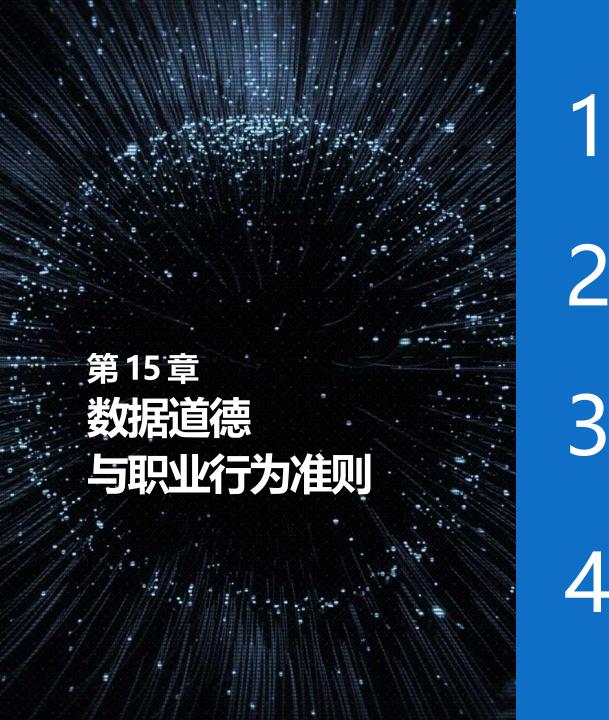
GitHub流行的原因





短短几年 Android/iOS开发者 的需求迅速下降

前端开发技术的演化



开放的世界

数据科学与工程职业规划

数据隐私与社会问题

数据与人工智能伦理

每一个新技术的出现都是为了替代掉使用旧技术的人的劳动力!

要懂机器 能操纵机器



要理解开源能玩转开源

技术的步伐不可阻挡



理论基础与分析能力

算法分析、逻辑推理、 解析推导、概率统计演算



工程能力

实现正确、高效的代码, 软件架构,前端,debug

互联、开放、协作、演化 促进新一代卓越IT人才的养成!



拥抱互联网与云计算,开放分享和协作学习, 成为具有开源精神的新一代卓越IT人才



• 沟通能力

读(论文), 写(技术文档和论文), 说(演讲,技术汇报,电梯talk)



创新与产品思维

对产品需求敏锐 利用技术创造出有用的产品

数据科学家能力体系

企业 开源社区 一

Information technology

Organization & Community

Education technology

Big Data

Cloud computing

Blockchain

Artificial intelligence



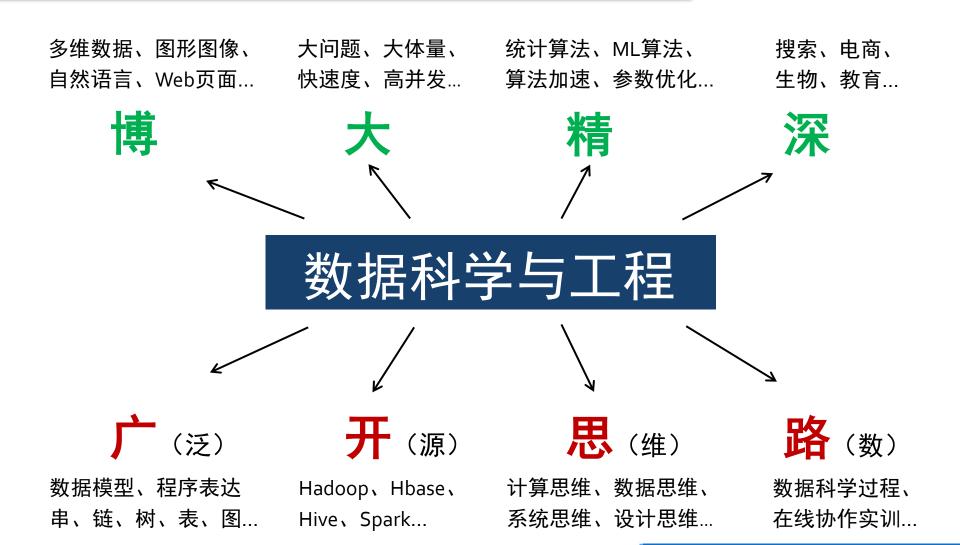
MOOX+新工科

教育信息科学与技术

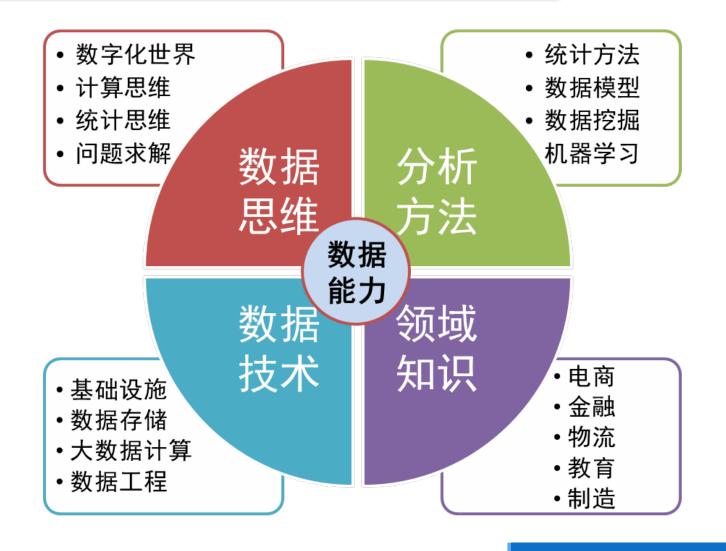
人工智能/CS4AII教育

知识经济/共享经济

校企社结合



再看"数据科学与工程"



数据能力模型

	数据思维	分析方法	数据技术	领域知识
复合型人才	***	**	**	****
咨询型人才	****	***	***	***
技术型人才	**	**	****	**
兴趣型人才	*	*	*	*

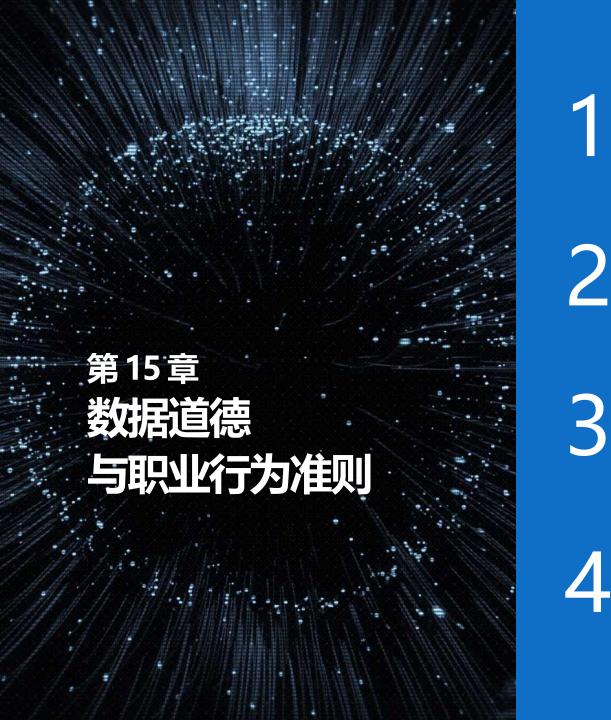
• <mark>思维模型</mark>:计算思维、数据思维、问题求解

• <u>分析方法</u>: 概率统计、数据挖掘、机器学习

• 数据技术:基础设施、数据存储、大数据、数据工程

• <u>领域知识</u>: 电商、金融、物流、教育、制造

数据人才知识矩阵



开放的世界

数据科学与工程职业规划

数据隐私与社会问题

数据与人工智能伦理





你该怎么做?

电车难题



保护车外的行人? 还是保护车内的乘客?



该保护谁?



推送广告 影响大选





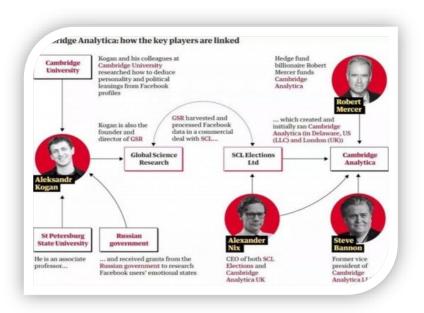


扎克伯格公开道歉

泄露用户数据

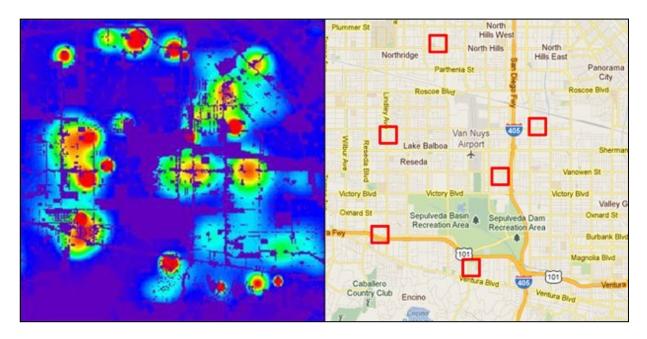


- 心理学 + 大数据 = 颠覆世界
- ・看懂一个人只需要68个赞



Cambridge Analytica

点击"Like"就能决定谁是总统?



算法生成的犯罪活动热度图

Automated Inference on Criminality using Face Images

Xiaolin Wu Shanghai Jiao Tong University

xwu510@gmail.com

Abstract

We study, for the first time, automated inference on criminality based solely on still face images. Via supervised machine learning, we build four classifiers (logistic regression, KNN, SVM, CNN) using facial images of 1856 real persons controlled for race, gender, age and facial expressions, nearly half of whom were convicted criminals, for discriminating between criminals and non-criminals. All four classifiers perform consistently well and produce evidence for the validity of automated face-induced inference on crimXi Zhang Shanghai Jiao Tong University

zhangxi_19930818@sjtu.edu.cn

people share the belief that the face alone suffices to reveal innate traits of a person. Aristotle in his famous work Prior Analytics asserted, "It is possible to infer character from features, if it is granted that the body and the soul are changed together by the natural affections". Psychologists have known, for as long as a millennium, the human tendency of inferring innate traits and social attributes (e.g., the trustworthiness, dominance) of a person from his/her facial appearance, and a robust consensus of individuals' inferences . These are the facts found through numerous studies [2, 32, 4, 5, 9, 20, 21, 27, 25].

基于人脸的自动犯罪性推测

是否赞成: 算法自动推测犯罪?

你是犯罪嫌疑人!



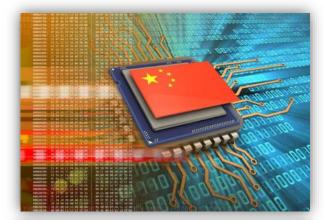
是否赞成:人工智能取代世界上90%的工作?

真正"解放生产力"!

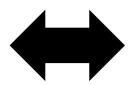
- 关于两篇文章:
 - 周志华, 关于强人工智能, 计算机学会通讯, 2018年1月.
 - 黄铁军, 也谈强人工智能, , 计算机学会通讯, 2018年2月.
- 给出你对强人工智能的看法。
 - 弱人工智能: 借鉴人类的智能行为, 研制出更好的工具以减轻人类智力劳动。
 - 强人工智能: 研制出达到甚至超越人类智慧水平的人造物, 具有心智和意识、能根据自己的意图开展行动。

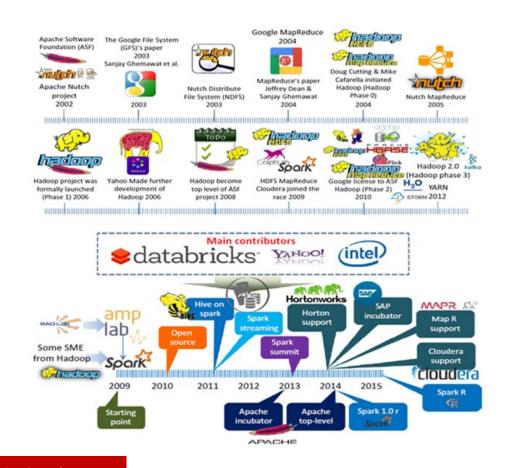
是否赞成:强人工智能优先弱人工智能?

让人工智能拥有思维?



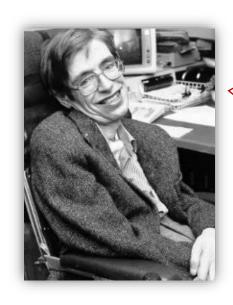






自主可控还是开放创新?

另起炉灶 or 协作共赢?



拥有智慧的外星 人,在发展自己 科技的同时,可 能也和地球人一 样过度消耗资源 ,最终走向毁灭 T.

Ted Kaczynski, 智商167, 16 岁哈佛, 25岁最年轻教 授, 却成为"大学炸弹客" 新技术的最大问题,就是 剥夺人类的自由。自由与 技术进步不相容,技术越 进步,自由越后退。

《工业社会和它的未来》

https://en.wikipedia.org/wiki/Ted_Kaczynski

- 科学,是开蒙人的,解开世界的谜,并带领人类探索未知的世界。
- 技术完全控制地球上的一切,人类自由基本将不复存在,因为个人无法对抗用超级技术武装起来的大型组织。只有极少数人握有真正的权力,但就连他们的自由也是十分有限的,因为他们的行为也是受到管制的。

技术决定论?



开放的世界

数据科学与工程职业规划

3 数据隐私与社会问题

数据与人工智能伦理

- 计算专业人员的行为改变世界。他们应 反思其工作的广泛影响, 始终如一支持 公众利益,才能负责任地行事。
- 旨在激励和指导包括现有和胸怀抱负的 从业者、教师、学生、影响者以及任何 以有影响力的方式使用计算技术的人士 等所有计算专业人员的道德行为。



that cannot be obtained elsewhere. Explain why anonymity was granted.

▶ Diligently seek subjects of news coverage to allow them to respond to criti-

· Avoid undercover or other surreptitious methods of gathering information

➤ Be vigilant and courageous about holding those with power accountable.

· Recoming a previal obligation to serve as wearthfood over redule affairs and

Provide access to source material when it is relevant and appropriate.

▶ Avoid storeotyping. Journalists should examine the ways their values and

➤ Never deliberately distort facts or context, including visual information.

Ohical journalism treats sources, subjects, colleagues and members of

cism or allegations of wrongdoing.

open, and that public records are open to all.

Seek sources whose voices we seldom hear.

Clearly label illustrations and re-enactments.

the public as human beings deserving of respect.

experiences may shape their reporting.

Never placiarize. Always attribute.

► Label advocacy and commentary.

Journalists should:

Give voice to the voiceless.

ACT INDEPENDENTLY

The highest and primary obligation of ethical journalism is to serve the public.

Journalists should:

- . Ausid conflicts of interest, real or perceived. Disclose unavoidable conflicts. unless traditional, open methods will not yield information vital to the public. . Refuse gifts, favors, fees, free travel and special treatment, and avoid political and other outside activities that may compremise integrity or importiality. or may damage credibility. · Support the open and civil exchange of views, even views they find repugnant.
 - . Be wary of sources offering information for favors or money, do not pay for
 - . Dany favored treatment to advertisers, donors or any other special interests.
- and resist internal and external pressure to influence coverage ▶ Boidly tell the story of the diversity and magnitude of the human experience. · Clistinguish news from advertising and shun hybrids that blur the lines between the two. Prominently label sponsored content.

BE ACCOUNTABLE AND TRANSPARENT

Dition journalism means taking responsibility for one's work and

Journellists should:

- · Exclain ethical choices and processes to audiences. Encourage a civil dialogue with the public about journalistic practices, coverage and news
- · Respond quickly to questions about accuracy, clarity and fairness.
- . Acknowledge mistakes and correct them promptly and prominently Explain corrections and clarifications cansfully and clearly.
- ➤ Balance the public's need for information against potential harm or disconflux. ➤ Expose unethical conduct in journalism, including within their organizations . Abide by the same high standards they expect of others.

read as a whole, individual principles should not be taken out of centest. It is not, our can it be under the First Amendment, legally entire sable

计算机协会道德与职业行为准则

15.4 数据与人工智能伦理

- 为社会和人类的幸福做出贡献,承认所有人都是计算的利益相关者。
- 避免伤害。
- 诚实可靠。
- 做事公平,采取行动无歧视。
- 尊重需要产生新想法、新发明、创造性作品和计算工件的工作。
- 尊重隐私。
- 尊重保密协议。

一般道德原则

15.4 数据与人工智能伦理

- 努力在专业工作的过程和产品中实现高质量。
- 保持高标准的专业能力、行为和道德实践。
- 了解并尊重与专业工作相关的现有规则。
- 接受并提供适当的专业审查。
- 对计算机系统及其影响进行全面彻底的评估,包括分析可能的风险。
- 仅在能力范围内开展工作。
- 培养公众对计算、相关技术及其后果的认识和理解。
- 仅当获得授权或仅为公众利益之目的才能访问计算和通信资源。
- 设计和实施具有稳固又可用的安全的系统。

职业责任



开放的世界

数据科学与工程职业规划

数据隐私与社会问题

数据与人工智能伦理