



# AI驱动的脑科学研究：数据共享、 开放科学与数据安全的平衡

演讲人：王安宇

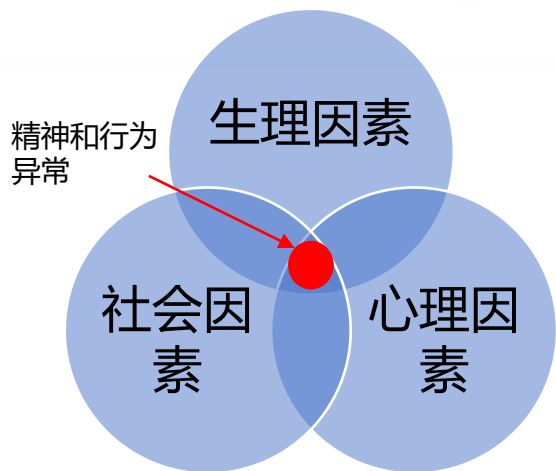
单位名称：天桥脑科学研究院 (TCCI)



- 01 AI驱动的脑科学研究
- 02 脑科学数据集建设
- 03 AI+脑科学的挑战与展望
- 04 关于TCCI

01

# AI驱动的脑科学研究



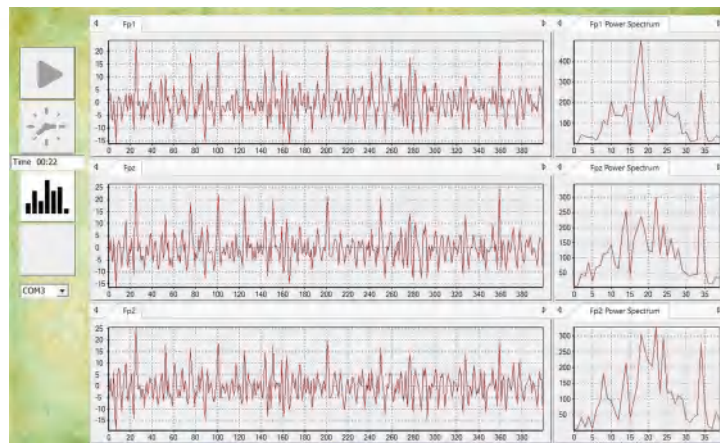
WHO 报告 (2001)

## 精神健康:

精神疾病是一种改变一个人的**思想、情绪或行为**（或兼而有之）的健康状况。

抑郁症、精神分裂症、注意力缺陷多动障碍 (ADHD) 和自闭症谱系障碍 (ASD) 等多发。

全世界约有 **4.5 亿人** 有精神健康问题。



## 诊疗难度:

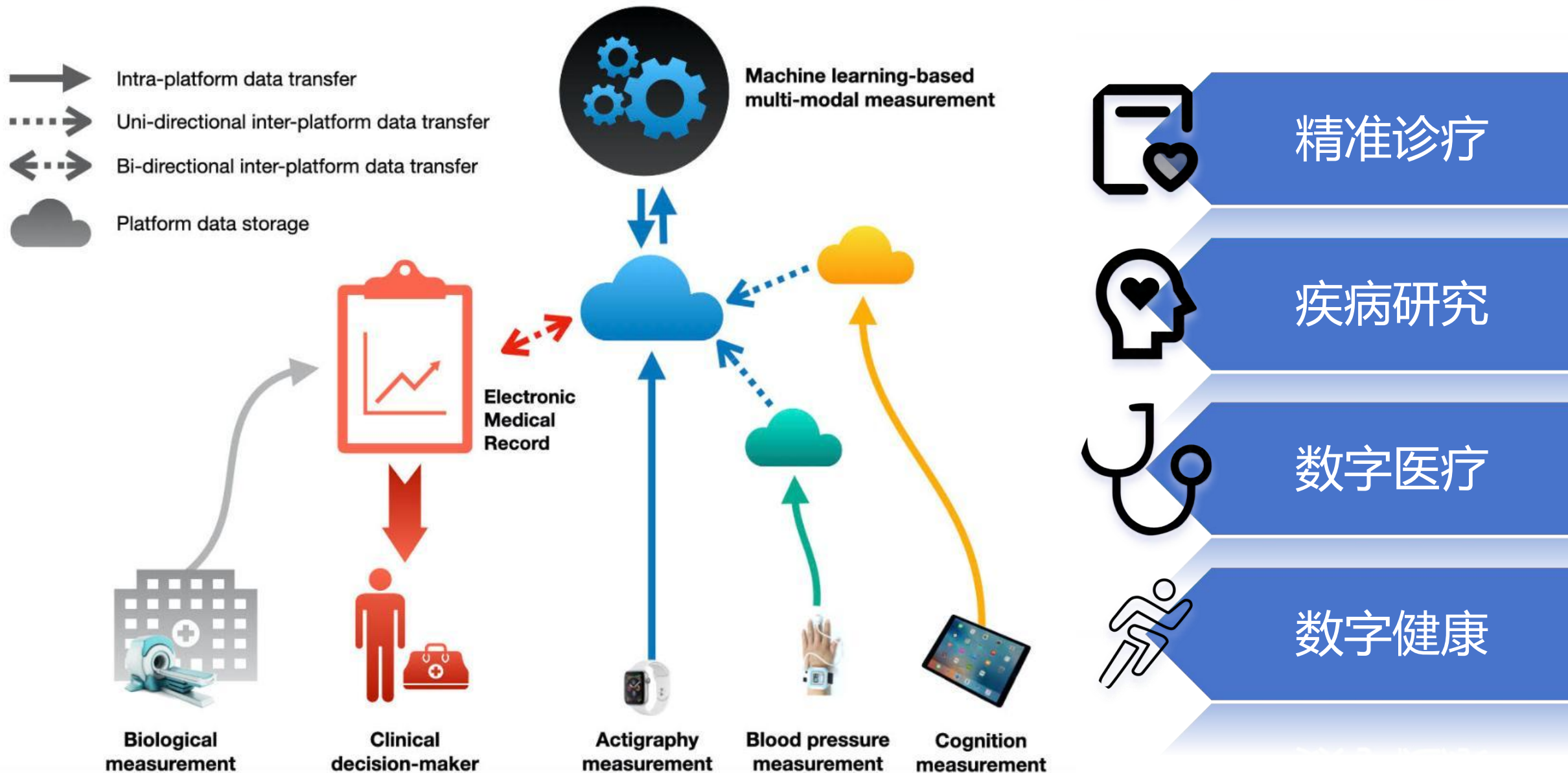
缺乏针对精神疾病的**生物标记**。无法依赖于医疗实验室测试和**指标判断**。精神疾病通常根据个人对特定问卷的自我报告进行诊断，这些问卷旨在检测特定的情绪模式或社交互动。

## 社会影响:

根据耶鲁大学经济学家阿列赫·茨温斯基的一项新研究，精神疾病每年给美国经济造成 **2820 亿美元** 的损失。美国卫生部(NIH)指出，2019 年，美国患有 MDD（重度抑郁症）的成年人数量估计为 **1980 万**，造成的增量社会经济负担估计为 **3337 亿美元**。

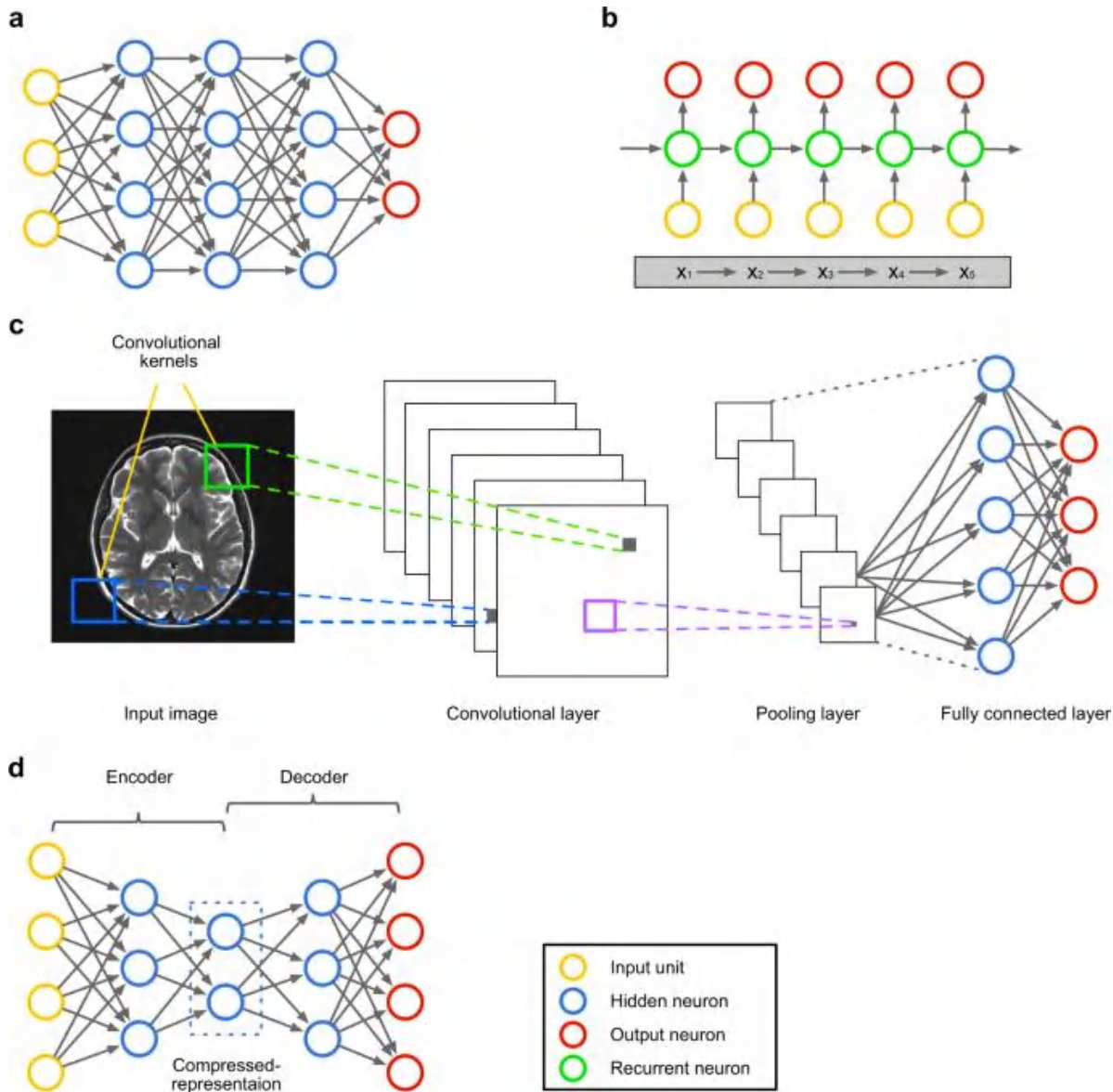


# 数+智，在健康医疗中的应用愈加广泛



ML 类型	概念	代表性方法	应用场景
<b>监督学习</b>	从标记数据中学习预测类别/临床指标	SVM、随机森林、稀疏学习、集成学习	疾病诊断、预后、治疗结果预测
<b>无监督学习</b>	从未标记的数据中学习以揭示结构并识别子组	层次聚类、K 均值、PCA、CCA	疾病亚型、规范建模、识别行为和神经生物学维度
<b>半监督学习</b>	从标记和未标记的数据中学习以执行监督或无监督的任务	多视角学习、拉普拉斯正则化、半监督聚类	多模态分析、联合疾病分型与诊断、不完整数据预测
<b>深度学习</b>	学习层次结构和特征的非线性映射以获得更高级别的表示，可以是监督的，也可以是无监督的	CNN、深度自动编码器、GCN、RNN、LSTM、GAN	一大类通用的学习问题
<b>强化学习</b>	解决时间信用分配问题、最优控制、反复试验学习	时间差分学习、Q 学习、演员-评论家模型、动态规划	在线控制、决策和选择行为建模

# 不同类型的“神经网络”

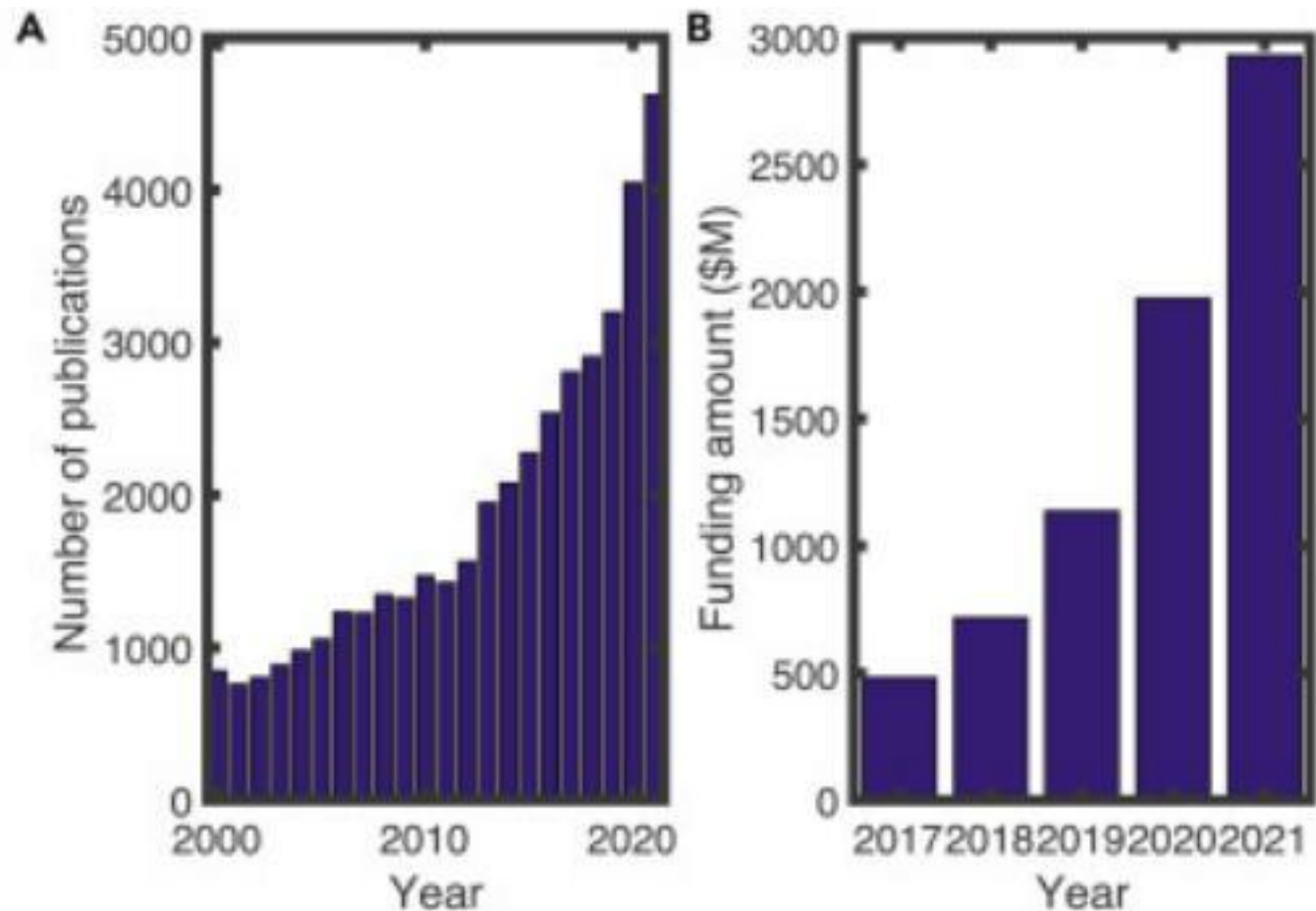


**a 深度前馈神经网络 (DFNN)**: DL 模型的基本设计。DFNN 通常包含多个隐藏层。

**b 循环神经网络 (RNN)**: 处理序列数据。对历史信息进行编码，每个循环神经元接收前一个神经元的输入元素和状态向量，并产生一个隐藏状态，馈送到后继神经元。

**c 卷积神经网络 (CNN)**: 在输入层（例如，输入神经图像）和输出层之间，CNN 通常包含三种类型的层：卷积层、池化层、全连接层。

**d 自动编码器**: 由两个组件组成：编码器，它学习逐层将输入数据压缩为潜在表示；而解码器（与编码器相反）学习在输出层重建数据。

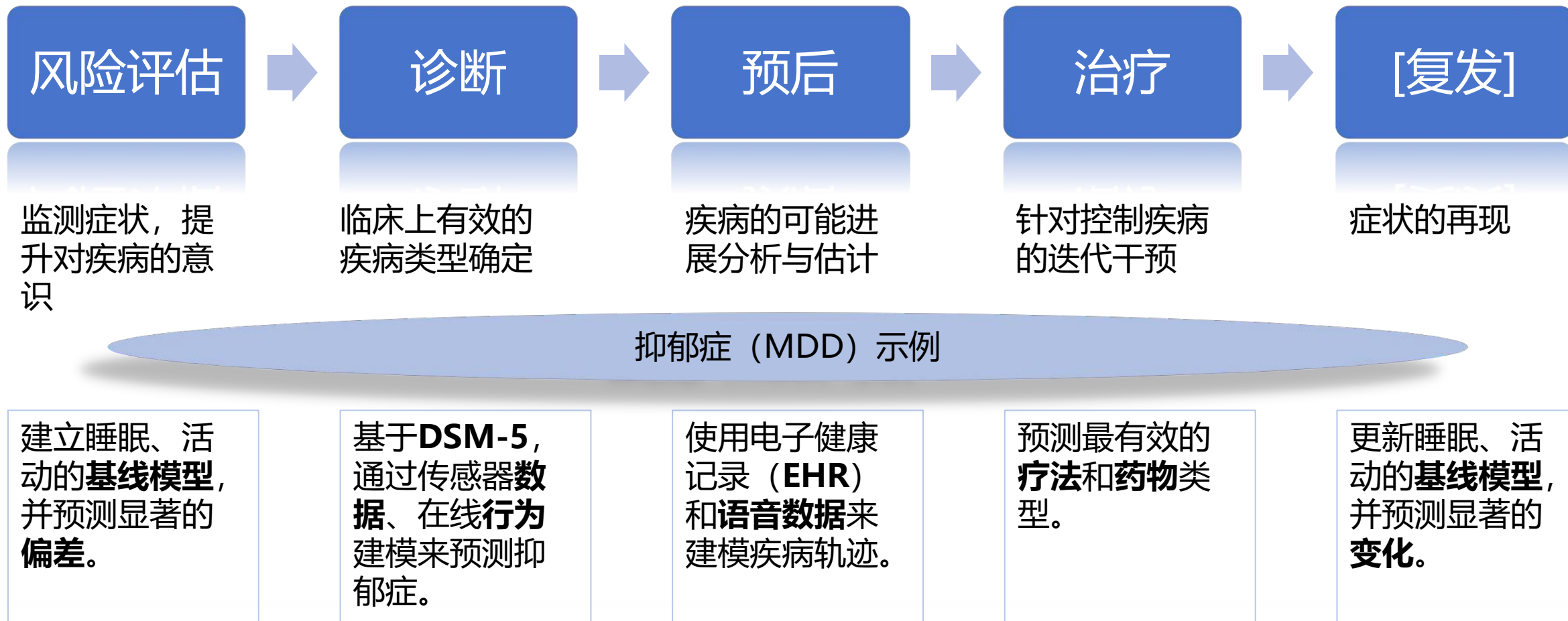


过去二十年，人们对将 ML/AI 应用于精神病学的兴趣稳步增长，这反映在 PubMed 出版物的数量上（图1 A）。

(B) 美国市场心理健康科技资金增长（2017-2021 年；数据来源：<https://www.cbinsights.com>）。



# AI 可应用于精神疾病诊疗的各个场景





# 02

# 脑科学数据集建设



Human Brain Project

10 Years of BRAIN  
A Decade of  
Innovation



## 欧盟：

2013~2023，总投资10亿欧元。  
显著的成就包括领先的数字大脑图谱、跨尺度的先进大脑模拟平台、认知模型和个性化医疗的应用，以及神经形态计算、神经启发机器人和人工智能方面的显著进步。

## 美国：

2013年发起，截止2022年，总投资24亿美元。  
旨在解析860亿个神经细胞及其彼此间形成的数以万亿计的连接。已确定了小鼠、狨猴以及人类的主要运动皮质区的100多种细胞类型。

## 日本：

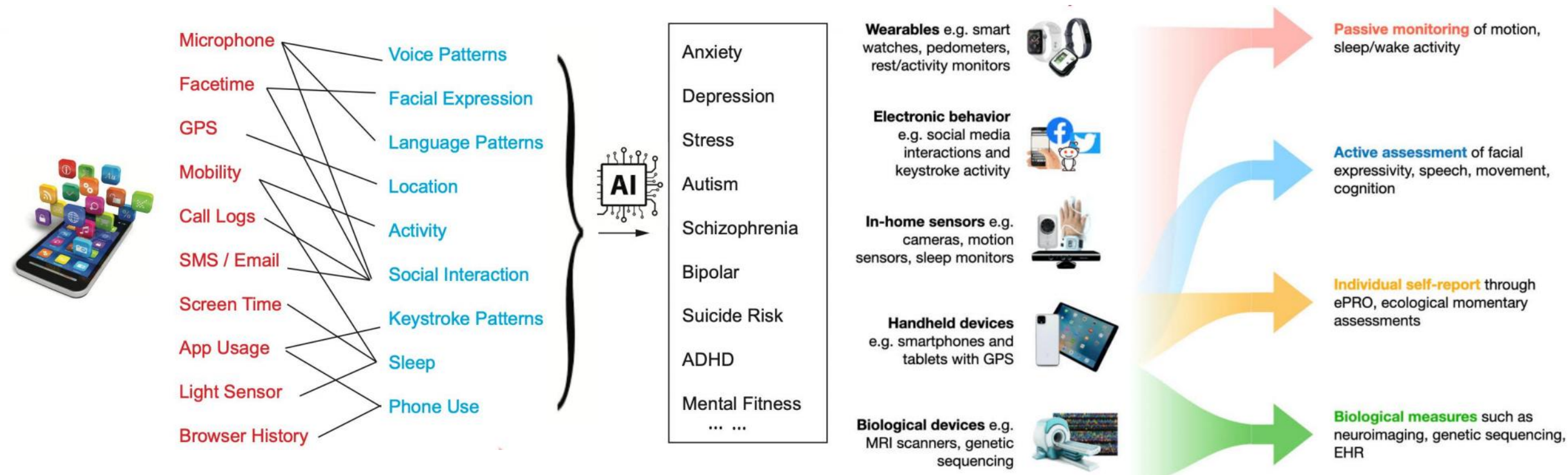
2018年启动。总投入3.65亿美元。  
主要围绕5个方面开展工作：发现和干预初期的神经疾病、分析从健康状态到患病状态的大脑图像、开发基于AI的脑科学技术、比较人类和灵长类动物的神经环路、划分脑结构功能区并开展同源性研究。



2021年9月正式启动，  
国家拨款经费预算近  
32亿元。

中国脑计划以“**脑认知功能解析**”为核心，以“**理解脑、修复脑、模拟脑**”为目标，确定了“**一体两翼**”的发展战略。

# 高质量数据集对实现精准医疗至关重要



## 被动监测

动作、睡眠、醒来

## 主动评估

面部表情、语言、行为、认知

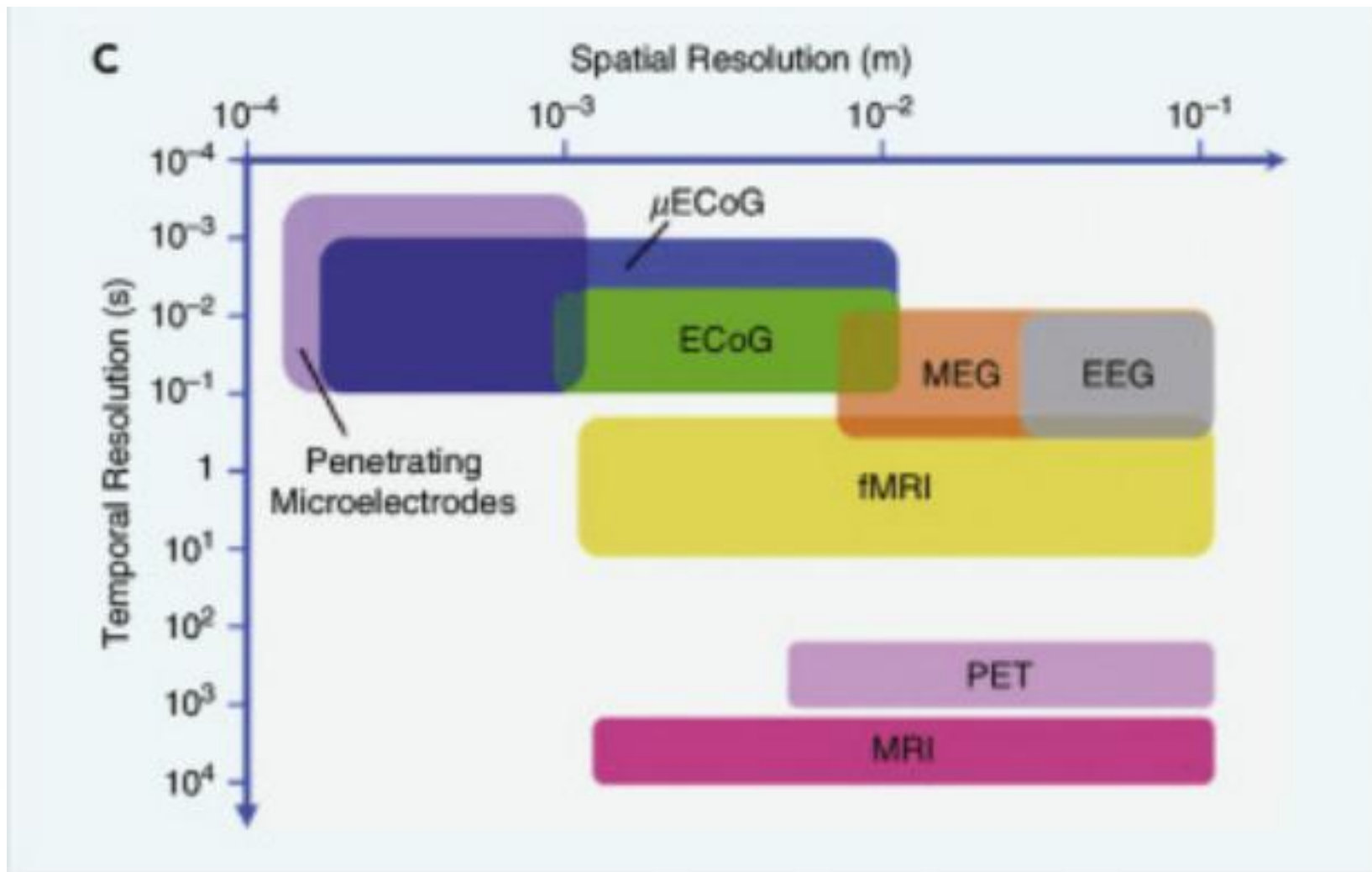
## 个体自报告

ePRO, 生态瞬时评估

## 生物测量

神经成像, 基因测序、EHR

**多模态数据集的兴起:** 影像等生物学数据、可穿戴设备收集的生理数据、社交数据、行为数据。



- 电生理记录 (electrophysiological recording) :
  - 脑电图 (EEG)
  - 皮层电图 (ECoG)
- 光学成像:
  - 双光子钙调蛋白成像
  - 电压敏感染料
- 功能性核磁共振成像 (fMRI)
- 功能性近红外光谱 (fNIRS)
- 脑磁图 (MEG)
- 正电子发射型计算机断层显像 (PET)
- 化学探针

各种神经记录技术，具备不同的时间与空间分辨率。

## Biobank



### 50万样本数据集（英国）

涵盖影像、基因、电子健康记录、生物标记、活动监测、问卷、生物样本。  
截止2023年12月，**9869**项论文已发表。

## ENIGMA



### 5万被试者数据集（国际）

成立于2009年，是一个世界范围的联盟组织。汇聚了来自35个国家的研究者。  
以多模态脑影像数据为主。

## OpenNeuro



### 1210份公开数据集（美国为主）

由 OpenfMRI 项目演进。涵盖**50737**位参与者，BIDS兼容的 MRI, PET, MEG, EEG, iEEG 数据。



推动国际合作和数据开放



推动科学进步



有利于人类福祉

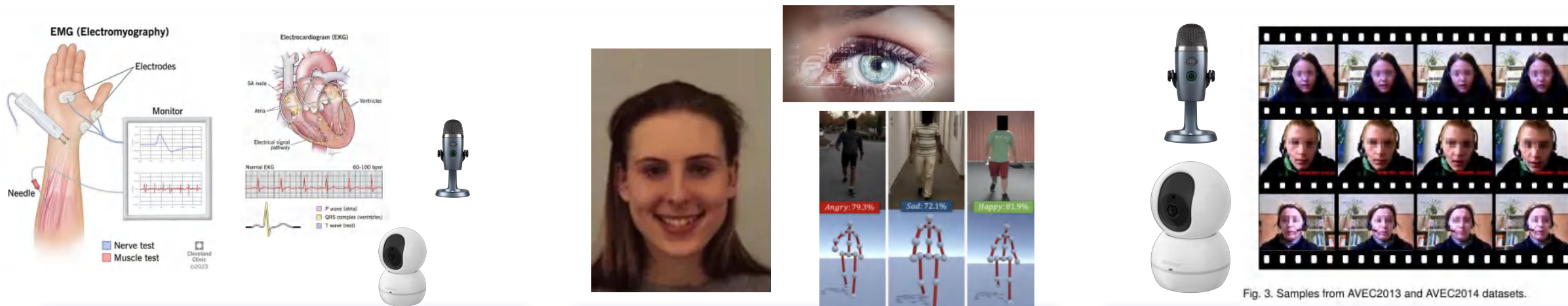


Fig. 3. Samples from AVEC2013 and AVEC2014 datasets.

## 多模态数据集:

生理指标 (肌电图EMG、心电图ECG、皮肤电导率SC和呼吸变化RSP等), 视频、音频等。  
MIT-BIH, Aabt, SAVEE等公开数据集。

## 视频情感数据集:

面部视频、眼动数据、步态等行动姿态。  
LIRIS-ACCEDE、DEAP、HUMAINE等公开数据集。

## 语音识别抑郁数据集:

通过声音语调等信息进行抑郁与控制组的分类预测。  
Mundt-35, AVEC-2013, AVEC-2014等公开数据集。

**仍存在样本量小 (几十到几百参与者), 使用频率不高、影响力不够大等不足。**





# 模型和数据集的开放，有利于科研合作



研究者共享数据、共享模型，通过比赛优化模型表现。

Kaggle Datasets search results for 'Mental Health':

- Covid-19 Lockdown Survey of Kolkata Residents** (Nhabhra Rohan Das, Updated 13 days ago, Usability 8.8 - 2 Files (CSV, other) - 50 KB)
- Mental Health Dataset** (Bhavik Jkadar, Updated 2 months ago, Usability 10.0 - 1 File (CSV) - 2 MB)
- Student Mental health** (MD Sharful Islam, Updated a year ago, Usability 8.2 - 1 File (CSV) - 2 KB)
- Behavioral Risk Factor Data: Tobacco** (Sutham Sarkar, Updated 2 months ago, Usability 10.0 - 1 File (CSV) - 1 MB)

Hugging Face Models search results for 'mental health':

Model Name	Updated	Views
Asod/mental_health_counseling_conversatio...	Updated Apr 5, 2023	3.6K
heliosbrahma/mental_health_chatbot_dataset	Updated Mar 1, 2023	316
solonok/reddit_mental_health_posts	Updated Jun 11, 2023	13
heliosbrahma/mental_health_conversational...	Updated Jul 23, 2023	879
namikpandya/mental-health	Updated Jul 17, 2023	136
ZahrizhaIAI/mental_health_conversational...	Updated Aug 25, 2023	40
npingale/mental-health-chat-dataset	Updated Nov 6, 2023	39
tolu07/Mental_Health_FAQ	Updated Nov 28, 2023	7
sujoyc66/heliobranhas_mental_health_instr...	Updated Dec 27, 2023	1
TVRRaviteja/mental_health_counseling_conv...	Updated 23 days ago	9
jsfactory/mental_health_reddit_posts	Updated Dec 11, 2021	7
alexandreteles/mental-health-conversation...	Updated Dec 28, 2022	161
quocanh34/mental_health_dataset_1	Updated Jun 12, 2023	4
PrinceAyush/Mental_Health_conv	Updated Aug 4, 2023	3
Riyazmk/mentalhealth	Updated Aug 5, 2023	4
fridriik/mental-health-arg-post-quarantin...	Updated Aug 26, 2023	1
open-lln-leaderboard/details_vibhorag101_...		
open-lln-leaderboard/details_Marshvir_La...		

和鲸社区 | 首页 | 频道 | 项目 | 数据集 | 比赛 | 活动

项目 3 | 数据集 6 | 比赛 0 | 活动 0 | 频道 0 | 专栏 0 | 用户 0

综合排序



### 关于全球心理健康调查数据集的探索

项目

可一键运行复现的 Notebook 数据分析项目

每周挑战

中华白海豚41hv 5 巡航长 · 2个月前 · 608 5 13 5 1



### 数据分析|Pyecharts:音乐和心理健康影响分析

项目

音乐和心理健康数据可视化

医疗健康

Secret Base 6 参谋官 · 9个月前 · 1.2k 3 21 12 1



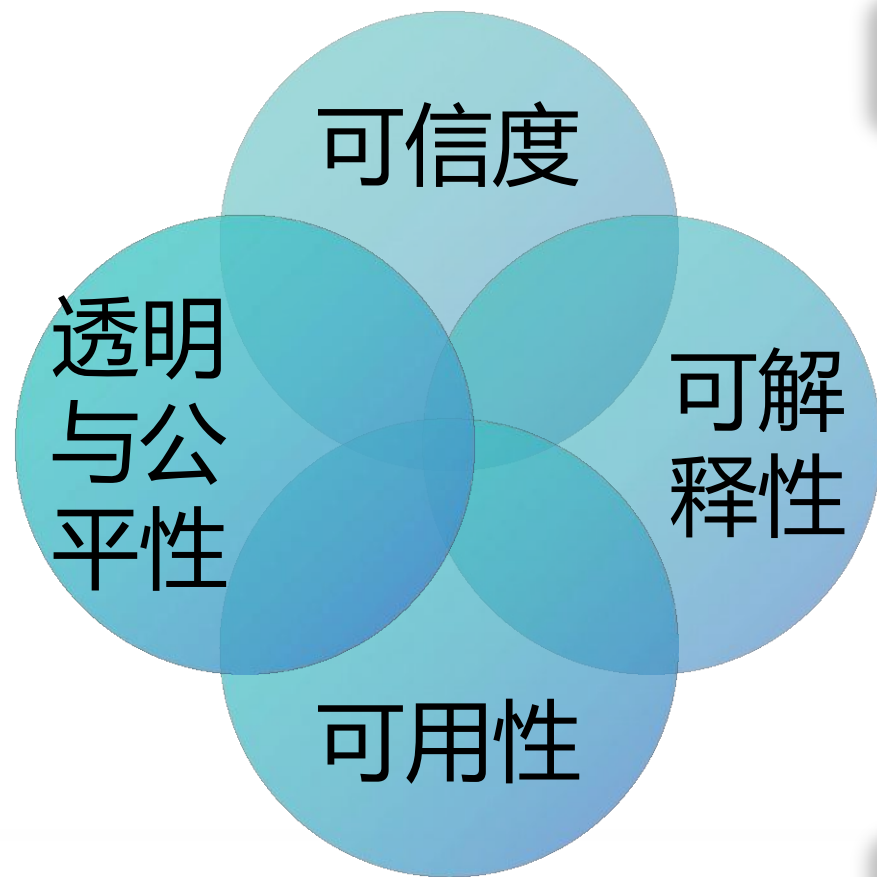
### MeChat: 中文心理健康支持对话大模型

项目

教你如何使用中文心理健康支持对话大模型，以及如何用自己的数据进行微调

# 03

# AI+脑科学的挑战与展望



可信度

评估 MI 得出的输出在不同输入和环境下的有效性和可靠性的能力。

可解释性

有权了解和理解数据集/输入的哪些方面可能会影响算法的输出（临床决策支持）。

透明度

以人类的视角理解和评估机器、算法或计算过程的内部机制的能力。

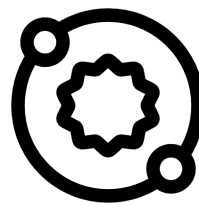
可用性

MI 系统在多种医疗环境中实现特定目标的有效性、效率和患者满意度的程度。

《自然·NPJ 数字医学》2020-47。



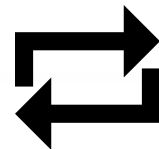
对数据及其适合度/质量/相关性进行“评分”。第一步需要能够从**组成、出处、代表性和完整性**方面描述数据集。



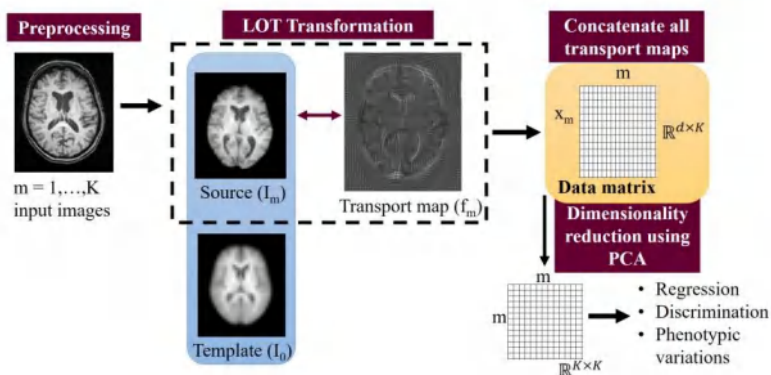
MI 系统的技术和概念方面的**可重复性和稳健性**（例如，针对对抗性示例）。



附加系统（可以是人）必须评估 MI 系统**输出的可信度**。

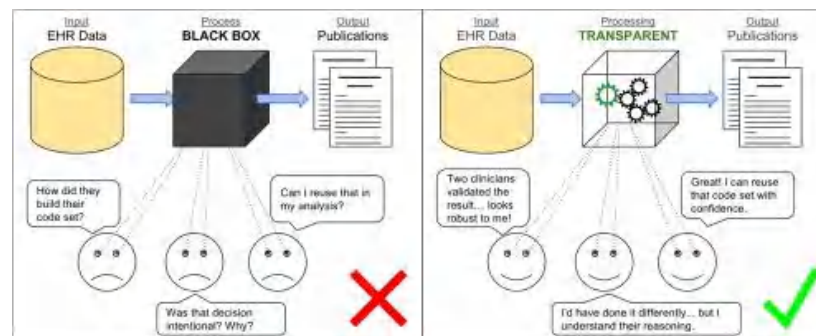


就像在其他行业一样，需要监控长期结果并建立结构化的**反馈链路**。



## 对数据处理的解释

示例包括生成图像变换、代理系统、遮挡可视化、显著图和类激活图。



## 提高输入的可见性

应用技术和最佳实践，表明系统如何确定输出。

## Review

the beer was n't what i expected, and i'm not sure it's "true to style", but i thought it was delicious. **a very pleasant ruby red-amber color** with a relatively brilliant finish, but a limited amount of carbonation, from the look of it. aroma is what i think an amber ale should be - a nice blend of caramel and happiness bound together.

## Ratings

**Look: 5 stars**

Smell: 4 stars

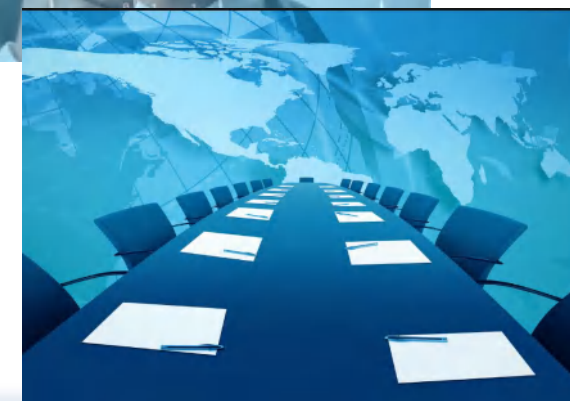
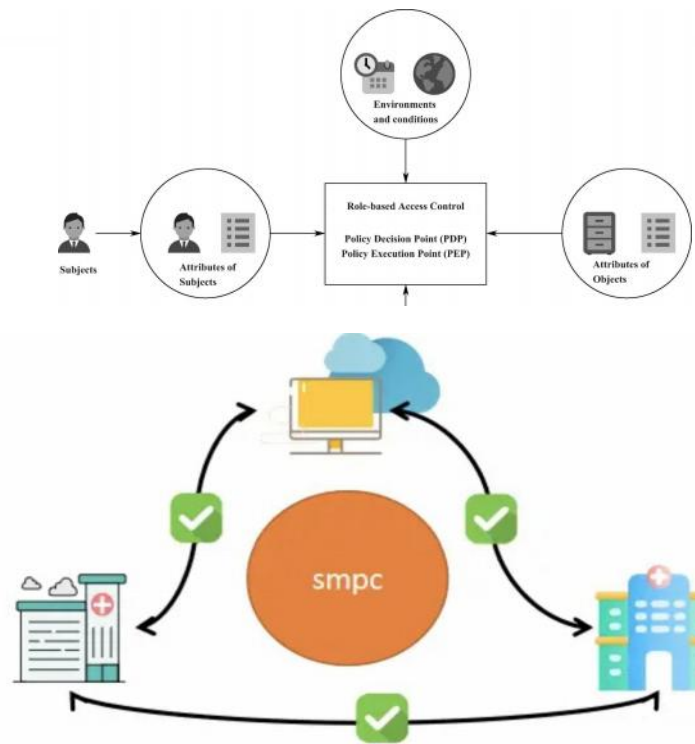
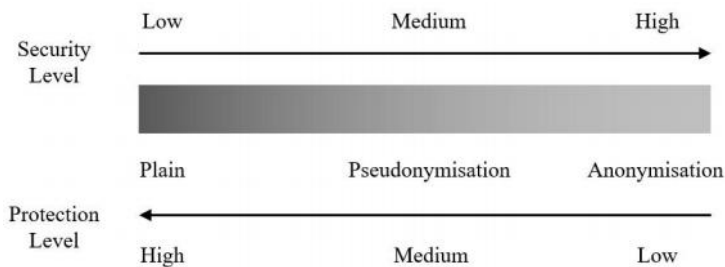
## 解释接口

将输出定量打分并可视化。

## 解释生成系统

通过网络解剖、解缠表示或明确训练网络来创建解释生成系统以生成解释。

Figure 1: An example of a review with ranking in two categories. The rationale for Look prediction is shown in bold.



## 数据匿名化和差分隐私:

降低数据关联到个人的风险。

## 零信任和多方计算:

确保敏感数据的传输安全和使用安全。  
仅限于授权人员和授权实体使用。

## 透明度与伦理审查:

通过伦理委员会和透明度声明确保研究的正当性。

## 开放科学的潮流



- ✓ 开放获取、开放数据、开放源代码、开放同行评审、开放实验
- ✓ 欧盟《开放科学政策》

## 跨境数据监管合作



- ✓ 脑科学数据共享
- ✓ 政策协调

## 增强公众信任



- ✓ 数据使用透明度
- ✓ 数据保护能力认同
- ✓ 数据使用效益认同

# 04 关于TCCI



# 全球最大的私人脑科学研究机构之一



## 天桥脑科学研究院 (Tianqiao and Chrissy Chen Institute, TCCI)

由盛大网络创始人陈天桥、雒芊芊夫妇私人出资10亿美元组建

聚焦全球化、跨领域和青年科学家



# AI驱动科学大奖

- 研究院与《科学》杂志共同发起“**AI驱动科学大奖**”，旨在推动AI技术在基础科研中的应用，以AI加速推动科学创新。



Chen Institute & Science  
Prize for  
**AI Accelerated  
Research**

THANK YOU!