

# 深度信息技术（精品）专辑（第四期）

- 可解释性机器学习
- 区块链
- 人工智能
- 自动驾驶和无人驾驶
- AR/VR/XR
- 开源教学与实践

中国开源软件推进联盟（COPU）

二〇二一年八月二十九日

# 目录

## ■ 可解释性机器学习

陆首群：打破机器学习黑盒子实现可解释人工智能·····4

董琳、程海旭、Animesh Singh：可信任的人工智能——基于开源的人工智能可解释性探讨及案例分析·····7

## ■ 区块链

段永朝：区块链的社区价值····· 19

## ■ 人工智能

清华——工程院等：《人工智能发展报告 2020》····· 49

COPU：清华——工程院等发布“人工智能未来重点发展技术方面”的商榷····· 51

## ■ 自动驾驶和无人驾驶

编者的话<sup>1</sup>：国内外自动驾驶和无人驾驶····· 54

## ■ AR/VR/XR

刘寿永、邹鹏程：XR 创新与实践····· 57

## ■ 开源教学与实践

编者的话：MIT 开设的一门网络实验课程——操作系统工程··· 71

陈渝：Xv6 操作系统工程课程介绍····· 72

---

<sup>1</sup> COPU 编者为陈伟、鞠东颖。

# 人工智能

# 打破机器学习黑盒子实现可解释的人工智能

陆首群

2021. 5. 16

机器学习/深度学习是一种强大的数据分析工具，是弱人工智能的代表。但机器学习/深度学习也是有缺陷的，它本质上是一项暗箱技术或盲模型，其训练过程不可解释、不可控制。图灵奖得主、贝叶斯网络之父 Judea Pearl 早在 2018 年就指出，当前机器学习理论有局限性，完全以统计学或盲模型（即黑盒子）的方式运行，所以不能成为强人工智能的基础；人工智能大师 Yoshua Benio 谈到，近年来，以深度学习算法为代表的人工智能技术快速发展，在计算机视觉、语音识别、语义理解等领域实现了突破，但其算法并不完美，有待继续加强理论研究；图灵奖得主、算法大师约翰-霍普克罗夫在 2019 年指出，对深度学习这个黑盒子，人们知道它在学习，但不知它怎么学习，人类可能会在五年后大体读出深度学习的数学理论。COPU 于 2020 年 6 月指出，机器学习/深度学习必须克服其自身的缺陷，打破黑盒子痼疾，实现可解释的人工智能，建立可解释的机器学习模型。

如今，可解释性人工智能（XAI）已成为全球人工智能研发的亮点。

早在 2019 年 8 月，COPU 就提出研发 XAI 的任务，这在国内是最早提出的，在全球也是最早提出这个任务的少数机构之一。在国内，2020 年 12 月，沈向洋教授提出：“拥抱开源，我们现在最重要的事情是要做可解释的人工智能”。2021 年 1 月，姚期智院士提出：“机器学习算法缺乏可解释性，很多算法处于黑盒子状态，这项人工智能的技术瓶颈亟待突破”。

2020 年 6 月，COPU 主办《第 15 届开源中国开源世界高峰论坛》，邀请 IBM 副总裁 Todd Moore 在会上作“可信人工智能（反欺诈、可解释、公平性）”的报告，从此至今，COPU 已收到全球研发可解释性人工智能的跟帖 48 件。但由于全球人工智能技术（XAI）尚未完全成熟，在研发 XAI 算法时，专家对各道演算程序的理解和操作具有不确定性，最后评估还只能靠人工，所以 XAI 演算结果或算法可能有出入，致使可解释机器学习难以推广应用。为此，COPU 要求 IBM Todd Moore 和人工智能研究所的 CTO Animesh 对 XAI 举出具体案例并进行解析和说明，对我们提出的 8 个问题进行逐个解答：

① IBM 列出研发 XAI 的具体案例是什么？

② 选用下列哪种方法进入运算？

- 可直接解释（内在解释）
- 事后解释
- 全局（模型级）可解释性
- 局部（实例级）可解释性

③ 选择什么工具？

- 如：决策树、规划库、块择表等

④ 如何捕捉特征？

⑤ 如何建模？

⑥ 如何找到算法？

⑦ 如何进行评估？

⑧ 不但要导出本案例结果，还要使 XAI 在使用中确定是否能保持信任、公正、透明和可解释？！

# 可信任的人工智能—— 基于开源的人工智能可解释性探讨及案例分析

IBM 董琳 程海旭 Animesh Singh

## 一、可信任的人工智能

IBM 一直致力于解决 AI 的可信问题，AI 赋能的系统在改变我们生活的同时，也表现出一些缺陷，如存在偏见，缺乏解释能力，并且易受到对抗攻击。IBM 研究团队提出一套科学方法来帮助用户设计可信任的 AI 系统。这套方法主要围绕以下三个亟待解决的问题：

- **可解释性**

决策和建议易于被开发者和使用者理解。对于预测结果，特别是负面的预测，往往需要回答为什么，可解释往往意味着可以预测系统的行为，从而增加系统的可控性。

- **健壮性**

安全可靠，不易受到因篡改数据而引发的攻击。AI 系统与其他软件系统一样，需要抵御潜在的攻击，对抗样本等攻击会让 AI 系统产生与实际相悖的结果。

- **公平性**

使用不含偏差的数据和模型，确保所有群体能被公平对待。AI 系统的智能属性要求其遵循公平的社会准则，带有偏

差的训练数据可能带来包含偏见的预测结果,不公平的 AI 系统不仅适用性差,而且会带来负面的社会和法律效应。

如果 AI 系统公平但不能抵抗潜在的攻击,不会是一个可信的 AI 系统,如果 AI 系统安全健壮但不能被清楚地解释,也不是一个可信的 AI 系统,因此要构建可信的 AI 系统,这三条缺一不可。针对这三个问题,IBM 开源了三个相应的工具套件,以求帮助用户构建可信的 AI 系统。

- 可解释性工具套件 AIX360<sup>2</sup>: 从数据、特征、模型等角度,针对同一个问题,提供基于不同受众的多角度解释能力。
- 健壮性工具套件 ART<sup>3</sup>: 针对潜在的对抗样本攻击,提供从评估、攻击和防御的全套方法,在数据处理、模型构建训练以及预测服务各个阶段提高模型健壮性的能力。
- 公平性工具套件 AIF360<sup>4</sup>: 提供多种基于数据分析和模型预测分析的公平性评价指标,同时支持在机器学习各个阶段消除数据和模型偏差的全套方法。

为了满足企业对技术整合支持、保障及服务的需求,IBM 将这一系列技术和 IBM 多年的人工智能实践集成到 IBM Cloud Pak for Data 和 IBM Watson OpenScale 等产品中。

---

<sup>2</sup> <https://github.com/Trusted-AI/AIX360>

<sup>3</sup> <https://github.com/Trusted-AI/adversarial-robustness-toolbox>

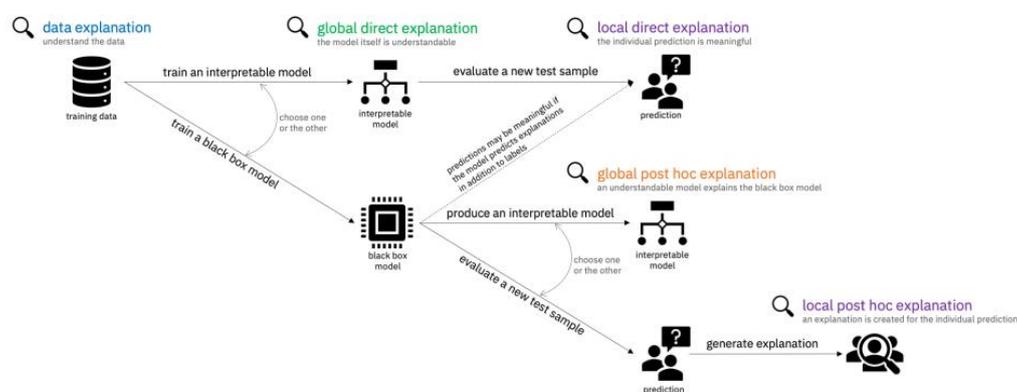
<sup>4</sup> <https://github.com/Trusted-AI/AIF360>

## 二、可解释的人工智能

### 1、技术背景

机器学习模型在越来越多的任务上表现出极高的准确率和广泛的普及。但是很多模型做出的推断并不能被使用者理解，只有理解了机器学习模型做出推断的过程，使用者才能信任基于模型的人工智能系统。

机器学习模型的最终用户可能会有各种各样的角色，每个角色会期望不同角度、不同程度的解释。借贷申请被驳回时，申请者会期望了解被驳回的原理，以及需要在哪些方面改进以得到批准。而监管者会更关注整个人工智能系统是否符合规范，而不是某一个具体推断的原因。开发者会关注模型的准确率、泛化能力等指标以不断优化模型。由此可见，当需要对机器学习模型的推断做出解释的时候，没有一个单一的方法，而需要根据机器学习的不同阶段和不同的使用者来提供不同关的解释方法。



IBM 开源的 AIX360 项目通过一套统一的接口，使用不同的算法，提供多维度的可解释性。

## 2、解释数据

机器学习开始于数据，了解数据的属性和特征能够帮助数据科学家有效地理解要解决的问题，进而采取合适的解决方法。有些数据的特征很容易理解，有些特征则因为包含了多重含义而会难以理解，DIP-VAE 通过无监督学习的方法根据已有的数据特征得到新的更容易理解的数据特征。而 ProtoDash 则通过从数据集中快速提取具有代表的样本原型来理解不同类别样本的具体差异。

## 3、解释模型

数据之后就是构建模型，最终执行推断任务的也是模型，一般可以有多种方式来解释机器学习的模型，这些方式针对的场景或者面向的受众会有所区别。第一个是直接可解释和事后解释的区别，直接可解释模型通常如同白盒一般，容易被人所理解，如决策树，线性回归等，而事后解释则是考虑如何解释针对黑盒模型构建或提供更易于理解的方法。第二个是全局解释和局部解释的区别，全局解释关注模型整体的可解释，而局部解释则更关注某个具体样本推断的可解释。全局直接可解释模型对于需要了解整个决策过程并确保其安全性、可靠性或合规性的角色很重，譬如监管人员和数据科学家。全局事后解释对于以来机器学习模型辅助做出决定

的角色很重要，譬如医生和贷款员等。局部可解释则对于收到机器学习推断直接影响的角色很重要，如果病人和银行客户等，这些角色更关系与其相关的决策的可解释。AIX360 包含了直接与事后解释、全局和局部解释的各种算法。

- **全局直接可解释模型**

AIX360 提供的 BRCG 和 GLRM 都是全局直接可解释模型，两个都适用于分类问题，GLRM 则还适用于回归问题。对于分类问题，BRCG 适合用来构建简单易于理解的模型，而 GLRM 则能构建准确率更高的模型，同时保持作为线性模型的可解释性。

- **全局事后解释**

AIX360 提供的 ProfWeight 算法，可以基于不可解释的神经网络黑盒模型学习得到训练样本权重，以此可以构建更加简单易于理解的模型。

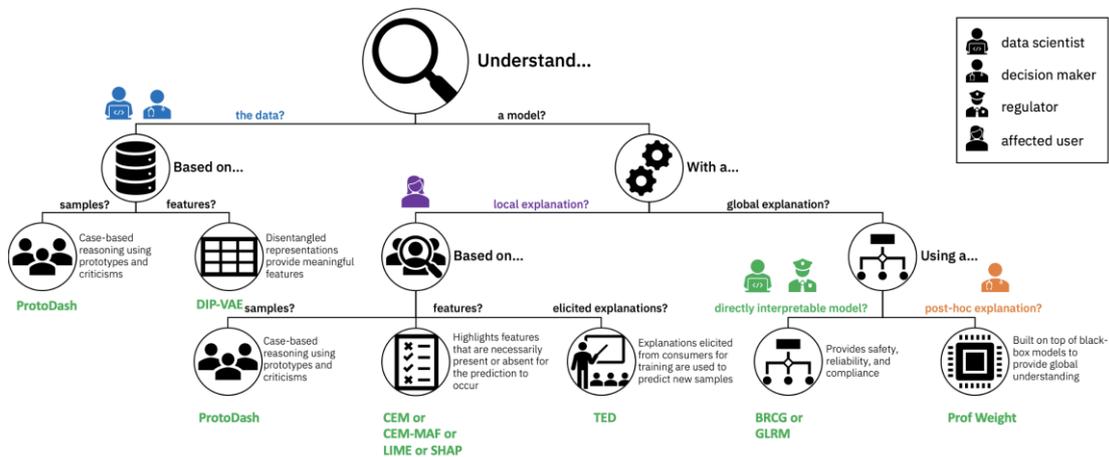
- **局部直接可解释模型**

AIX360 提供的 TED 可以直接训练一个在样本级别可解释的模型，其主要的思路是在训练数据中除了常见的推断标签之外，再提供一个解释标签，如此模型在预测时，也可以直接给出相应的解释。

- **局部事后解释**

AIX360 提供的 ProtoDash 和 CEM 等算法可以为每个样本的解释做出解释，ProtoDash 通过在历史数据中找出与推断

样本类似的原型并对比原型的标签和特征，以此解释对于该样本的推断，而 CEM 可以通过分析样本的特征，直接给出维持当前推断的主要特征，而改变该推断所缺少的特征。



### 三、可解释性人工智能在银行业务中的应用

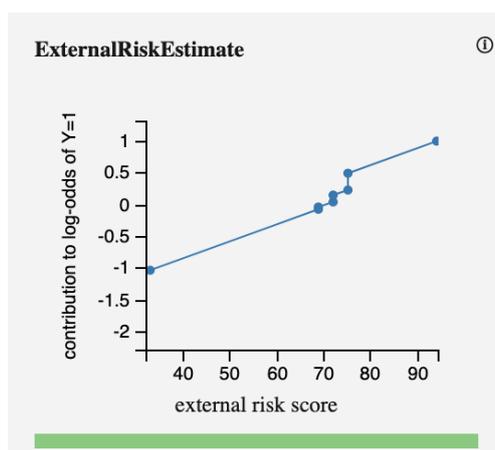
一家信用评级公司需要基于用户的房屋净值信用额度数据来判断用户能否在未来两年及时还付申请的贷款，以及决定是否批准贷款。贷款员将通过机器学习模型来判断是否可以批贷以及额度。AIX360 将为该业务中三个不同的角色（数据科学家、贷款员和银行客户）提供不同角度的解释<sup>5</sup>。

#### 1、数据科学家

对于数据科学家来说，他更期望从模型的整体上来理解模型的推断过程，而不是某个具体的贷款申请。AIX360 提供的 Generalized Linear Rule Model (GLRM) 算法是全局直接可解释模型，可以组合各种影响贷款的条件，以此推断申请

<sup>5</sup> <https://aix360.mybluemix.net/data>

人及时还款的可能性。同时 GLRM 可以通过训练学习的到样本中各个特征（条件）对于还款风险的影响程度，如下图所示。通过查看各个特征的影响程度，就能够个很容易理解 GLRM 赋予各个条件的组合权重。



## 2、贷款员

贷款员是最终决定用户申请批准与否的人，他们期望理解机器学习模型推断的具体原理，以此来做错正确且理由充分的审批。AIX360 提供的 ProtoDash 算法可以通过在历史样本中寻找与当前申请记录相似的样本，通过比较历史相似样本的特征和结果，可以验证机器学习模型对于当前申请的推断。如下图所示，当对 Robert 的申请给出拒绝的时候，通过 ProtoDash 在历史记录中找到个跟其最相似的记录结果也是拒绝。

	Robert	James	Danielle	Franklin
Outcome	-	Defaulted	Defaulted	Defaulted
Similarity to Robert (from 0 to 1)	-	0.690	0.114	0.108
ExternalRiskEstimate	78	71	72	69
MSinceOldestTradeOpen	82	95	166	193
MSinceMostRecentTradeOpen	5	1	12	12
AverageMInFile	54	43	74	167
NumSatisfactoryTrades	33	33	37	36
NumTrades60Ever2DerogPubRec	0	0	1	0
NumTrades90Ever2DerogPubRec	0	0	1	0
PercentTradesNeverDelq	100	100	95	100
MSinceMostRecentDelq	0	0	7	0
MaxDelq2PublicRecLast12M	7	7	4	7
MaxDelqEver	8	8	4	8
NumTotalTrades	41	41	41	8
NumTradesOpeninLast12M	2	4	0	0
PercentInstallTrades	15	17	15	6
MSinceMostRecentInqexcl7days	0	0	0	0
NumInqLast6M	3	4	1	0
NumInqLast6Mexcl7days	3	4	1	0
NetFractionRevolvingBurden	21	17	16	85
NetFractionInstallBurden	11	89	0	0
NumRevolvingTradesWBalance	9	7	3	16
NumInstallTradesWBalance	3	3	1	0
NumBank2NatlTradesWHighUtilization	2	1	1	13
PercentTradesWBalance	50	53	26	71

为便于理解，对表中各行特征的含义做进一步说明。

特征	含义
ExternalRiskEstimate	综合风险标记
MSinceOldestTradeOpen	最早账目的时长（以月为单位）
MSinceMostRecentTradeOpen	最新账目的时长（以月为单位）
AverageMInFile	账目的平均时长（以月为单位）
NumSatisfactoryTrades	合规账目数量
NumTrades60Ever2DerogPubRec	拖欠超过 60 天以上的账目数量
NumTrades90Ever2DerogPubRec	拖欠超过 90 天以上的账目数量
PercentTradesNeverDelq	未拖欠账目占比

MSinceMostRecentDelq	最近一次拖欠账目距今的月数
MaxDelq2PublicRecLast12M	过去 12 个月内最差拖欠分数
MaxDelqEver	最差拖欠分数
NumTotalTrades	总账目数量
NumTradesOpeninLast12M	过去 12 月账目数量
PercentInstallTrades	分期付款账目占比
MSinceMostRecentInqexcl7days	距离 7 天前最近一次信用查询的月数
NumInqLast6M	近 6 月信用查询次数
NumInqLast6Mexcl7days	近 6 月信用查询次数（不包含最近七天）
NetFractionRevolvingBurden	循环债务余额占信用额度的百分比
NetFractionInstallBurden	分期付款债务余额占原始贷款金额的百分比
NumRevolvingTradesWBalance	含余额循环债务账目数量
NumInstallTradesWBalance	含余额分期付款债务账目数量
NumBank2NatlTradesWHighUtilization	高利用率账目数量
PercentTradesWBalance	含余额债务账目比例

### 3、银行客户

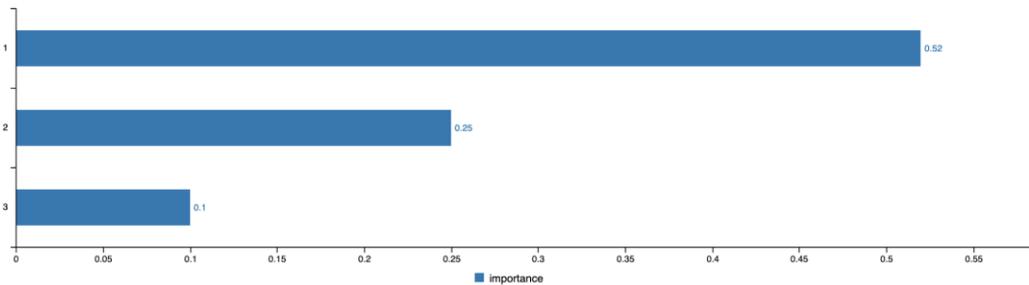
银行客户作为贷款申请人，他们期望知道申请被通过和拒绝的原因，特别是在被拒绝的情况下。AIX360 提供的 Contrastive Explanations Method (CEM) 算法可以发现用户申请记录的特征中可以改进的项，以及需要改进到什么程度可以通过审批。

如下图中当申请被拒绝时，系统将给出导致该结果的主要原因：

- 1) 综合风险标记为 77，需要该值为 82 及以上时才可能被批准。
- 2) 合规账目数量为 10，需要该值为 13 及以上时才可能被批准。

3) 过去 12 个月内最差拖欠分数为 6，需要该值为 7 及以上才可能被批准。

同时给出这几个主要原因对该推断结果的影响程度，可以看到综合风险标记对于结果的影响最大，客户可以在首先该方面进一步增强。



## 四、可信 AI 的 8 个原则

Linux 基金会 Data & AI 提出了构建可信任 AI 系统的 8 个原则<sup>6</sup>，除了上文提到的可解释性、健壮性和公平性外，还有可重现性，隐秘性，负责性，透明性和安全性。这些原则相互依赖和影响，共同作用以构建可信任的 AI 系统。

**可重现性：**在同样的环境下，不同的团队，利用相同的数据、框架以及算法等可以得到相同的结果，以保证 AI 系统提供的结果可靠，而不是因为某种巧合或不为人知的规则。

**隐秘性：**保护 AI 系统构建和使用过程中各个步骤相关资产的隐私，如训练数据、算法模型和推断数据等。

---

<sup>6</sup> <https://lfaidata.foundation/blog/2021/02/08/lfa-ai-data-announces-principles-for-trusted-ai/>

**负责性:** AI 系统和构建 AI 系统的人应该对 AI 系统做出的决定或行为进行解释、证明和承担责任,保持对 AI 系统变化的追溯和行为的监控。

**透明性:** 充分说明 AI 系统工作的细节,包括其提供的能力和潜在的不足,保证其工作流程和推断的透明。

**安全性:** AI 系统应该被深入测试和验证,以确保在明确定义的领域里发挥作用,并保护受 AI 系统影响的群体。

## 五、结语

构建可信任的 AI 系统是 AI 落地极其重要的一步,IBM 联合开源社区已经做出了一些卓有成效的探索,并在某些产品中提供了相应的能力。但可信任是一个需要持续探索和不断投入的领域,在此也欢迎更多有志于构建可信任的 AI 系统的人能加入到这项工作中来。

# 区块链

# 重塑：区块链的社会价值<sup>7</sup>

段永朝

2020-5-20

区块链严格说并非是一项独立的技术，它是若干基础技术的综合，是“汇聚技术（Converging Technologies）”的典范。区块链的基础技术包括：密码学、Hash 函数、分布式网络、可信计算、对等网络等等。“汇聚技术”是技术嵌入生命的重要历史现象，或者用海德格尔的术语，是生命嵌入技术“座架（Ge-stell）”的过程<sup>8</sup>。这一历史现象在第一次工业革命广泛使用碳能源、蒸汽机械之后，就开启了一个日益加速的历史进程。

在技术汇聚日益导致“人类增强”的大背景下，区块链技术应运而生，并迅速成为世界各国竞相关注的战略领域，充分显示区块链已成为自 1969 年互联网诞生以来，为数不多的几项重新塑造未来世界形态的伟大技术之一。这一伟大技术的清单可以从不同的角度去罗列，但至少会包括包交换技术、分布式计算、万维网技术、人工智能技术等。

2016 年之后，世界主要经济体的政府首脑、央行行长纷纷关注区块链，并相继将区块链纳入国家战略。技术专业媒

---

<sup>7</sup> 本文是段永朝老师关于区块链的一篇文章，部分内容摘要发表在《香港经济评论》等媒体。

<sup>8</sup> “技术、道术与家：海德格尔批判现代技术本质的意义及局限”，张祥龙，《现代哲学》，2016 年第 5 期，p.59

体和大众媒介，纷纷加入解读区块链时代意义的行列，这在技术演化史、传播史中并不鲜见。从 150 年前的电灯、电话，到 100 年前的汽车、飞机，以及近百年来的广播电视与电脑互联网，这些重大的技术创新，都伴随着人类日常生活、生产方式、社会组织方式的塑造。1920 年，列宁对共产主义的诠释，就是这样一个简单的公式：共产主义=苏维埃+电气化<sup>9</sup>。技术与社会形态的共生关系，一再提示人们在一项新的技术涌现之后，除了需要密切关注其技术动向、经济价值之外，还需要深入思考其社会含义和文化意味。区块链作为一项汇聚技术，它将对未来带来何种巨大变革？它的意义到底何在呢？这里试图从三个关键词：信任、共识、知足社会，来粗略地探讨区块链的社会价值和未来影响。

## 1. 信任：从多次重复博弈到陌生人的信任

信任问题无疑是人类发展中的重要问题。德国哲学家雅斯贝尔斯将其总结为“轴心时代”的核心问题。在雅斯贝尔斯《论历史的起源与目标》中，他将 2500 年前左右，发生在北纬 25 度到 35 度区间内的文明形态，称为“轴心时代”<sup>10</sup>。轴心时代面临的根本挑战，是如何建立世界秩序的问题。古希腊、希伯来、古印度和华夏四大文明区域的贤哲，总结出这样一个建构秩序的“金规则 (Golden Rules)”<sup>11</sup>，可视为影

---

<sup>9</sup> “人民委员会的工作报告”，1920 年，《列宁论重工业的发展和全国电气化》，人民出版社 1956，p.108

<sup>10</sup> 《论历史的起源与目标》，【德】雅斯贝尔斯，华东师范大学出版社 2018，p.7

<sup>11</sup> 《轴心时代》，【英】凯伦·阿姆斯特朗，海南出版社 2010，p.4

响至今的信任法则。这一法则的中文表述是“己所不欲勿施于人”<sup>12</sup>；用《旧约圣经》的话说，则是“以眼还眼以牙还牙”<sup>13</sup>。

这种贯穿整个人类数千年文明史的“金规则”，在社会实践中有两个特征：其一是代价巨大，其二是冤冤相报。这两个特征的根本原因，是人与人之间建立信任的过程，是通过多次“打交道”的基础上形成的，所谓“吃亏上当长见识”、“日久见人心”。社会的基本信任是建立在多次重复博弈的基础之上。在建立信任、谋求共识的过程中，不停地有人投机取巧、抄近路、抖机灵，让遵守规则的人感受到了莫大的伤害和不公。美国纽约大学哲学教授詹姆斯·卡斯称之为“有限游戏”，即“零和博弈”<sup>14</sup>。

这种多次、重复的零和博弈，让经济生活、商业活动中的两大成本居高不下——缔约成本、履约成本。这种“交易成本”的存在，被经济学家科斯用来解释现代企业存在的理由<sup>15</sup>。因为信息不对称或者自利动机，买东西的时候要货比三家、讨价还价，从而产生大大小小的缔约成本。又因为“道德风险”、“劣币驱逐良币”带来的“柠檬市场”，经济学家发现“制度”也是维系自由市场运转的“经济要素”。然而，制

---

<sup>12</sup> 《论语：颜渊》第十二章

<sup>13</sup> 《旧约圣经：申命记》；又见《汉谟拉比法典》，转引自《巴比伦与亚述文明》，于殿利，商务印书馆2013，p296-304

<sup>14</sup> 《有限与无限的游戏：一个哲学家眼中的竞技世界》，【美】詹姆斯·卡斯，电子工业出版社2013，p.3

<sup>15</sup> “企业的性质”，科斯著，盛洪、陈郁译，《企业、市场与法律》，格致出版社，2009-6，p.130

度的建构又会进一步引发“公地悖论”，带来更多的履约成本。为什么我们一定要在降低交易成本、化解道德风险方面花费那么多的时间、精力、资源？这是生产、交易、分配和日常生活绕不过去的“坎儿”吗？重提这一问题，对深入理解区块链背后的思想，很有必要。

在区块链思想出现之前，众多社会学家、经济学家，甚至生物学家，都试图从各自领域解释或者解决“合作”的问题。我们可以举三个例子。第一个来自美国密歇根大学的政治学教授罗伯特·阿克塞尔罗德（Robert Axelrod），他在1980-1984年间组织了两场计算机竞赛游戏，他抛出的问题是：获利的最佳策略依然是弱肉强食，但是能不能在自私的个体之间达成合作呢？最终多伦多大学的一个团队获胜，他们采取的游戏策略就是“己所不欲勿施于人”。在阿克塞尔罗德的《合作的进化》（The Evolution of Cooperation）一书中，提出人的天性中不仅包括自私的一面，也存在合作的一面。

哈佛大学的数学与生物学教授马丁·诺瓦克（Martin A. Nowak）也在数学、生物学领域上探索，认为合作是继突变和自然选择之后的第三个进化原则。诺瓦克研究物种的群体行为，研究物种和物种彼此之间的互惠行为。他发现挠痒痒和互相梳毛，是上百万年生物演化过程中遗存下来非常重要的亲社会行为。

纽约大学宗教历史系教授詹姆斯·卡斯(James P. Carse)在《有限与无限的游戏》(Finite and Infinite Games)中认为,过去2000多年来人类是在“有限游戏”里,其特征就是资源竞争、论输赢、有生死,这种残酷竞争的意识被社会达尔文主义者推波助澜,变成一种社会思潮。詹姆斯·卡斯认为,人类社会演进过程中的各种博弈可以分成两类:一类是有限的,必有终结,必有输赢;一类是无限的,没有终结,比如母子之间的互动,必须信任和互惠才能有和谐共赢的长久关系。卡斯说,“未来的游戏是无限游戏”,不会因为局部输赢而整个游戏就被彻底“玩坏了”。

2016年,谷歌公司的智能下棋机器人AlphaGo战胜李世石等围棋高手之后,很多棋手和围棋爱好者一开始觉得很受挫,认为在具体的对弈中,人终将无法战胜机器,“作为竞赛的围棋”已经毫无悬念地完败机器人。甚至有人认为,围棋已经索然无味。但这只是问题的一面。AlphaGo战胜人类棋手,只是“有限游戏”。但对于人而言,“下棋”的乐趣依然存在,AlphaGo让我们重新理解围棋,也重新理解人与机器、人与围棋的关系。

在信息时代,数据成为新的生产要素之后,新的分工法则(特别人与机器之间、机器与机器之间的分工法则,或者说“界面”)是什么,新的组织形态(特别自组织)是什么,成为信息经济、信息社会研究的迫切问题。在这些问题的背

后，如果建立信任的方式不发生改变，未来的信息社会将无法正常运转。为什么？因为我们联结的广度和深度远远超过传统社会。信息社会的一个显著特征，可以用“万物互联、虚实交互”这个术语来描述。当人们通过无数的传感器彼此相连，通过智能技术的“嵌入”成为赛博格（Cyborg）；当数字世界成为实体世界的“孪生世界”，每个人不但拥有生物学意义上的生命，进而拥有数字世界的多重化身的时候，人已经演变为新的物种，并接纳智能机器也成为新的“生命现象”。在这种未来画面之下，如果人与人之间建立信任的方式，依然停留在“多次重复博弈”的层级，停留在“零和博弈”的价值状态，这个世界的正常运转是不可想像的。

在万物互联的世界中，我们希望跟陌生人的信任一次就能够建立起来，我们希望能超越轴心时代所谓“金规则”。这听上去有乌托邦的意味，但在数字化进程方兴未艾的大背景下，已经充满了现实的可能性。我们可以从三个方面来观察：其一，是拓展“生命”的边界。18世纪的生命观带有强烈的机械论色彩。法国思想家拉-梅特里的小册子《人是机器》，就鲜明地表达了这种生命观。进入20世纪之后，特别过去100余年里，对大脑的研究，对DNA的研究，以及对意识的研究，在过去40年内形成认知科学、认知神经科学、系统生物学、社会生物学等诸多交叉学科，对“生命”的理解开拓了诸多新的边疆。试举几个重要的概念：涌现、自创生

等概念。

其二是拓展了“世界”的理念。“世界”是一个现代概念，伴随着西方民族国家兴起和工业资本主义进程，在其原初的含义中有浓厚的西方中心主义色彩。英国历史学家汤因比将文明纳入历史研究的视野，但他的世界观念依然是不同文明演进的物理综合。结构人类学的方法，打破了层次递进、单线条演进的史观，承认了文明进程的平行性和共存性。20世纪60年代，西方出现了新史学思潮，麦克尼尔等将人类历史的纬度，放大到作为一个物种的出现，即人的早期演化、迁徙。“人类世（Anthropocene）”概念的出现，将人的活动作为重要的“地质营力”，再次拓展了“世界”的纬度<sup>16</sup>。互联网、智能科技技术，进一步提出了新的可能。王飞跃将这一架构称之为“平行世界”，他认为，“未来世界的和谐，一定是‘物理世界 + 心理世界 + 人工世界’的和谐，加起来就是平行世界”<sup>17</sup>。数字资产研究院创办人、著名经济学家朱嘉明教授认为，“区块链首先是非物质世界和非物理世界的基础结构。其次，它是物质世界和非物质世界两个世界相连的桥梁。”<sup>18</sup>这一世界的观念演化，已经日益凝结为一股认知世界的洪流。“世界”这一概念的变迁，显示了历史意识、人类意

---

<sup>16</sup> “为什么要讨论‘人类世’？”，段永朝，《意义互联网：新轴心时代的认知重启》，东方出版社，2019-9，p.18

<sup>17</sup> “X5.0：平行时代的平行智能体系”，王飞跃，《中国计算机学会通讯》，2015，11(5)：p.10-14

<sup>18</sup> “关于形成区块链应用数学的可能性”，朱嘉明，区块链数学科学会议，2018年12月18日

识的变迁。

其三是拓展“账户”的概念。在全新的认知视野不断拓展的时代，可以说区块链思想的出现，一下子击中了人们的神经。透过这种建立在加密技术、分布式计算和对等网络架构基础之上的“账户”体系，为未来数字世界构建秩序、为未来数字生活和信息社会构建信任，奠定了坚实的基础。但要注意的是，这里的账户其实是“广义账户”的概念。

早在2014年，上海证券交易所前总工白硕就用“记账+认账”这四个字，非常简明扼要地概括了区块链技术的核心特征。<sup>19</sup>在这里，账户是区块链技术处理的重要对象。什么叫账户？简单说可以账户就是任何一笔经济活动的财务记录。所谓广义账户，就是说它不只是包含经济活动，它还包含社会活动。换句话说，在今天的互联生活中时刻发生的点赞、转发、评论、上传，都可以成为账户体系的一个记录。甚至未来数字孪生的世界中还可能包括情绪账户、心理账户，也就是说任何人有关其生存状态的全方位记录，都可以成为广义账户所记载的内容。这里我们暂且将相应的伦理、法律等问题放到一边，聚焦这个广义账户在缔结信任中的基础功能。

从狭义角度说，区块链支撑下的信任机制，建立在账户信息不可伪造、不可篡改、不可抵赖等方面。区块链通过加

---

<sup>19</sup> 《区块链真相》，武卿，机械工业出版社，2019-8，p.32

密手段确保账户信息的保密性和完整性。经济活动和社会生活一旦纳入区块链账户，永久不能赖账，财务造假的事就不存在了。但对于广义账户而言，区块链技术就不只是防止账户造假这个层级，它还有“信任评估”的功能。“日久见人心”这一古老的谚语，在区块链上得到了真实的体现。一个人的信用状态，不只通过具体的交易过程来体现，也综合了社会行为的点滴记录。数字世界的信用度量系统，简单说都是“攒积分”的过程。通过攒积分来积累社会声誉，并“威慑”违规行为。这有赖于社会行为数据与经济交易数据之间的互联互通。在这种广义账户的体系下，消费者意愿得到了充分的保证，个性化消费、个性化定价、个性化生产才真正成为可能。

从这个意义上说，区块链告诉我们，未来的信任的基础设施发生变化以后，人与人之间的经济关系从某种意义上说简单了。“漫天要价，就地还钱”永远成为过去了。价格作为“信号”的必要性大大降低。

综合上述，区块链技术使得社会信任的“立足之本”发生了根本性的变化，从“多次重复博弈”转向“陌生人一次信任”。工业时代的信任，是建立在信息不对称假设之上的；认为信息不对称无法逾越。信息时代，最大限度的压缩了信息不对称；换句话说，技术上完全可以做到信息对称。那么信任的基础就发生了变化。把信任建立在全新的基础上，靠

什么？靠的是区块链的“共识机制”。

## 2. 共识：从同意到计算

什么是共识机制？按照区块链的定义，共识机制是指通过特殊节点的投票，在很短的时间内完成对交易的验证和确认；对一笔交易，如果利益不相干的若干个节点能够达成共识，我们就可以认为全网对此也能够达成共识。主要的共识机制有工作量证明（POW）、权益证明（POS）等。从这个定义可以看到两个内涵：其一，区块链思想中的共识，是建立在计算的基础上的。工作量证明是算力证明，权益证明是权力证明；其二，区块链所谓的共识机制，就是前述白硕通俗解说的“认账”机制。某种程度上是传统共识概念的延伸。

这两个特点，可以归纳为一个，就是区块链所谓共识机制，是建立在“多数同意”的基础上的。这用技术术语说，区块链所谓的“共识”其实是分布式计算的一致性问题，其早先的技术思想，源自1959年兰德公司的Edmund Eisenberg和布朗大学的David Gale研究特定概率空间中，一组个体的主观概率分布如何形成“共识概率分布”的问题。在分布式计算研究领域，1975年，美国纽约州立大学石溪分校的E. A. Akkoyunlu和K. Ekanadham、R. V. Huber首次提出的虚拟问题“两军问题”，以及1980年Marshall Pease、Robert Shostak、Leslie Lamport提出的“拜占庭将军问题”。这些

技术思想成为区块链共识算法的工程基础。相关内容可参考中科院自动化所袁勇、王飞跃的《区块链理论与方法》一书的第四章“共识算法”<sup>20</sup>，这里不复赘述。

这里需要讨论的共识思想的社会学、政治学，乃至伦理学含义，并借此理解“共识”在区块链中的主要意义。值得注意的一个问题是：共识是可计算的吗？或者换一个说法，可计算是否是共识的一个基本特征？讨论这个问题，我们需要先看一看在社会科学领域，“共识”一语有什么含义，以及如何理解区块链背景下共识的意义。

“共识”一语的定义可谓五花八门。这里我们参考政治学里“共识”的定义：“在一定的时代，生活在一定的地理环境中的人们共有的一系列信念、价值观和规范准则。在政治意义上，它指的是与政治体系有关的信念。”<sup>21</sup>这个定义虽然是政治学意味，但其实更多强调的是信念、价值和准则这样的“道德”含义。

共识的哲学思考，源自英国哲学家洛克和休谟。洛克的共识理论可视为西方共识理论的开端，也是社会契约和古典自由主义的思想源泉。洛克致力于探讨如何为被宗教和道德分歧所分裂的国家，以及多远社会的出现提供共同的基础，以维护政治共同体的团结和统一。休谟将精神现象的动因或

---

<sup>20</sup> 《区块链理论与方法》，袁勇、王飞跃，清华大学出版社，2019年11月

<sup>21</sup> 《布莱克维尔政治学百科全书》，戴维·米勒，韦农·波格丹诺，中国政法大学出版社，2002年，p.166

最终原理，归结为“同情”，认为“同情”是一切道德感的来源，而共识的同一性来源于同情的稳定和同一。“同情”也被称作“同理心”。

对于启蒙运动的思想家来说，政治社会的秩序依赖于大众出于同意或者共识而对政治共同体给予持续的忠诚和支持。也就是说，公众对政治体系的支持要建立在普遍的道德根基之上，而不是依靠纯粹的利益基础，否则政治秩序就会崩溃。通过将政治共同体建立在道德共识的基础之上，洛克的共识理论，遂成为基于道德的共识。

洛克认为，理性是人类所具有的不同于动物的进行推理、演绎的能力，“它可以扩大我们的知识并调节我们的同意”。进而，洛克认为，“借着这种作用，人心就知道任何两个观念间的契合或相违，就如在解证中便是；它借此也看到概然的联系，就如在意见中便是。在前一方面，他就得到知识；在后一方面，它就予以同意或不同意。”<sup>22</sup>

洛克与霍布斯，都强调同意下的共识。这也是契约理论和人权理论的道德基础。只有同意才能提供人们组织政治共同体、经济共同体并将其视为合法的基础。“人类天生都是自由、平等和独立的，如不得本人的同意，不能把任何人置于这种状态之外，使受制于另一个人的政治权力。”<sup>23</sup>

---

<sup>22</sup> 《人类理解论》，【英】洛克，商务印书馆 1983，p.666

<sup>23</sup> 《政府论》，【英】洛克，商务印书馆，1982-11，p.57

洛克将共识与信任联系起来，共识缺乏的地方，社会信任和政治信任都是缺乏的，政治体系也是不稳定的。

建立在洛克道德基础上的共识理论，有两个假设：一个是作为个体的主体独立性假设；另一个则是个体同一性假设。前者是笛卡尔的“主客两分法”予以保证的；后者则是康德的物自体、黑格尔的绝对精神予以保证的。在高扬理性至上的17-19世纪，这种共识毫无疑问成为凝聚社会最大的道德力量。然而，情况并非如此简单。20世纪关于自由与平等的争论，关于民主与正义的争论，以及集权主义、自由主义的尖锐对抗，更不必说两次世界大战和持续半个世纪的意识形态冷战，都不无嘲讽地指出，这种建立在洛克道德哲学基础上的共识是多么的脆弱。

美国政治学家罗尔斯在谈论公平的正义之时，就试图抛开这种形而上学的方式，来探讨如何将政治共识建立在更加坚实的基础之上。他认为需要从道德哲学重回政治哲学，将公平正义的道德色彩，转移到政治意义下的公平正义，这样就可以促进正义观念在不同的政治理念之间获得最大限度的“重叠共识”。换句话说，罗尔斯的政治哲学，是建立在共识而非真理的基础之上的，他追求的是一致同意的共识，但完全可以保留对事实截然相反的判断。

罗尔斯的正义观是平等主义的，即一种用于社会基本结

构的权利、利益分配的正义观。他指出主要有两条理想选择的正义原则，即用于社会基本结构中的权利、利益分配的正义观。它主要有两条原则：一条是平等分配基本自由和在机会公平条件下，差异化分配社会经济利益的正义；另一条是这种分配原则要最可能地有利于最少获益者。<sup>24</sup>

公平的正义，是洛克重叠共识的内核。洛克将分配问题纳入公平原则，是开创性的。任何人不但有“有所为”的自由，但同时也必须对“所为之果如何承受”予以顾忌。也就是说，一个自由的人，不能肆无忌惮地使用自己的自由，而对自由之果无所顾忌。当然，洛克的重叠共识是否能够达成，是否有足够的逻辑基础和道德支撑，争议也是非常激烈的。比如诺齐克，就反对洛克的正义观。诺齐克认为，自由原则是至高无上的，分配正义没有道德基础。诺齐克认为，分配一词并非中性的，持有才是。“在一个自由的社会里，广泛而不同的人控制着各种资源，新的持有来自人们的自由交换和馈赠。”<sup>25</sup>诺齐克将罗尔斯的分配正义转换为持有正义，这又回到个体正义的原则，并将其作为自由社会的权利基础。

基于各种道德假设和政治意愿的争论，对什么是共识恐怕难以达成更加令人信服的共识。这说明什么问题？这可能说明这样一件事：传统共识的“谈论”方式，已经进入了死

---

<sup>24</sup> 《正义论》，【美】罗尔斯，中国社会科学出版社，2001-6

<sup>25</sup> 《无政府、国家与乌托邦》，【美】诺齐克，社科出版社 1991，p.150

胡同。前文所述轴心时代的“金规则”是两千多年来人们形成的“共识”，这一共识无论有多少种补充意见，都无法撼动其作为伦理学元命题的核心地位。然而，这一状况正在发生巨大的变化。变化的根源就在于我们正遭遇一个巨变的时代，这一时代是通过接近 50 亿网民、海量的数据、万物互联的可能性、以及智能装置的嵌入来表征的。

自苏格拉底以来，传统的西方学者将共识的达成，视为一个理性思辩、理性对话的过程。比如哈贝马斯坚持认为，完全合乎理性标准的对话，必定能产生一致认可的理解。中国学者赵汀阳指出，“哈贝马斯忽略了一个关键性问题，这就是，理解并不能保证接受。”“达成共识与合作的充分理由不是互相理解，而是互相接受”，互相接受超出了知识论和理性所能处理的范围。<sup>26</sup>

无论政治学者和社会学者对共识有多少种主张，有一点是他们共同的软肋，即共识如何度量？理解这一点，才能深入思考区块链共识的伟大意义。

换一个说法，就是共识是否可以计算的问题。上个世纪 90 年代以来，伴随信息技术的巨大挑战，国际哲学界提出了所谓“计算转向”的问题，这一转向对哲学、社会科学的影响透过近十年云计算、大数据、人工智能、虚拟现实、区块

---

<sup>26</sup> “论道德金规则的最佳可能方案”，赵汀阳，《中国社会科学》2005 年第 3 期，p.70

链的勃兴，将透过计算建构“共识”提上了议事日程。<sup>27</sup>中科院自动化所研究员王飞跃指出，“从技术本质而言，以AlphaGo 为代表的人工智能方法之意义可用一个‘AlphaGo Thesis’概之：AlphaGo 展示了从牛顿的‘大定律，小数据’技术范式向默顿的‘大数据，小定律’技术范式转移的可行性。”<sup>28</sup>

共识之所以遭遇挑战，是因为社会经济政治文化活动的频度和复杂度，节奏加快，对共识的衡量超出了人们依靠传统只是经验驾驭的程度，从而失却了价值直觉。思想解放的直接后果就是多元价值观的涌现。多元价值观的直接遭遇，使得价值冲突呈现悖论的场面。一方面多元价值主张的思想源头都可以追溯到彼此巧合的金规则；另一方面，附着在各自解释系统之下的价值伦理又异常庞杂，经常表现出尖锐对立，从而失去从容对话的共情基础。在这种情况下，一再地重申金规则的低阶共识，已经无法弥合巨大的认知差异和心灵震撼，共识升维的需要就自然涌现了。人类需要共识，金规则也不能轻率抛弃。但从哪里升维呢？可计算的共识或许是一条值得探究的出路。可计算的共识，正是今日区块链的意义所在。

区块链的共识机制，不仅包括传统共识理论所言的“共

---

<sup>27</sup> “信息哲学的兴起与发展”，刘刚，《社会科学管理与评论》2005年第2期，p.74

<sup>28</sup> “人工智能：第三轴心时代的来临”，王飞跃，《文化纵横》2017年12月，p.94-100

同认可的信仰、共同信守的准则”，也包括“接受结果”的思想和情感准备，同时将“共识”问题转化为“计算”问题。如何对待他人，最终要落实到如何采取行动上来。如何采取行动，又要转变对行动的认知。行动不是一劳永逸地，永久地改变，而是在变动中行动，在摸索中矫正或调整自己对待他人的方式，以保持长程视角，广域开放视角中的“金规则”。

这其实是将传统静态的金规则（有限游戏），转换为更高阶位的、动态的、相互嵌入、相互缠绕的金规则（无限游戏）。然而，随之而来的一个问题是，计算下的共识，是建立在“逻辑一致”、“道德共情”的基础上的吗？或者说，如何理解“计算下的共识”？

区块链的共识是建立在计算基础上的，区块链共识的计算性，又是纳入分布式网络的环境中的。密码学和分布式计算环境，使得共识的达成不是一个价值判断，更不是道德判断，而是“无法确知的大多数人”原则。这个原则看上去依然是“旧规则”，但在数字环境、网络环境之下，它具有了全新的特点：其一，“己所不欲勿施于人”的元命题，转换为“人所不欲勿施于人”（赵汀阳），增加对他者接受的关切；其二，网络环境理论上确保“全连接”空间成为超越个体的“对话空间”；其三，区块链加密特征确保行为数据沉淀为价值判断的基本要素；其四，区块链分叉机制，最大限度包容多元价值判断的可能性，生成平行区块链。王飞跃认为，区块链和

分布式账本本质上是一种新型的数据结构和系统架构，能够实现去中心化、去信任、安全可靠的描述智能；而未来以虚实结合、人机混合为特征的智能系统和复杂社会系统，更多地是需要基于预测智能 (Predictive Intelligence) 和引导智能 (Prescriptive Intelligence) 的实验、分析与决策<sup>29</sup>。这一思想延续了王飞跃 2004 年提出的 ACP 理论<sup>30</sup>。

可计算的区块链共识，将对未来的社会建构起到何种支撑作用？最近 5 年来，诸多区块链应用围绕数字货币、供应链金融、共享经济、知识产权等领域，展开了丰富多彩的探索。在笔者看来，这些探索无疑对加深区块链的社会意义有很好的启示。但值得注意的一个问题是，业已重新定义“共识”和“信任”的区块链，其社会价值仅仅是将传统工业社会的生产方式，转换为更高交易效率、更低交易成本的形态，仅仅是从技术角度确保“避免数据造假”、“可信信息追溯”吗？问题的焦点，恐怕需要回到这样一个基点：充分互联的社会环境下，社会财富增长的目的到底是什么？这就是下一个关键词“知足社会”。

### 3. 知足社会：从串行到并发

传统的农业社会和工业社会给人们带来的一个很大的

---

<sup>29</sup> Parallel Blockchain: Concept, Techniques and Applications, Wang FeiYue, The First International Symposium on Blockchain and Knowledge Automation, ISBKA 2017

<sup>30</sup> “人工社会、计算实验、平行系统: 关于复杂社会经济系统计算研究的讨论”, 王飞跃, 《复杂系统与复杂性科学》2004 年第 1 期, p.25-35

误区，就是哲学家杜威所说的“确定性的寻求”，或者叫定数崇拜。定数崇拜的根源是什么？就是杜威开篇所称的“人生活在危险的世界之中，便不得不寻求安全。”人类寻求安全的途径有二，一为祈祷、献祭、礼仪和巫祀，另一个则是艺术。通过这些对威胁其生命安全、决定命运的各种力量的崇拜和变相的“降服”，“人就从威胁着他的哪些条件和力量本身中构成了一座堡垒。”<sup>31</sup>

从定数崇拜的视角看，轴心时代人们达成共识的办法是彼此说服或者彼此征服。因此建立在轴心时代“金规则”基础上的共识，也可称之为“确定性共识”。

然而在互联网大行其道之后，共识这个事情发生了变化。变化的根由在于，越来越广泛的连接——包括人与人、人与机器、机器与机器，使得“确定性”不再是整个世界得以构建的基础。或者说，世界无法仰仗基于逻辑推断、道德共情、政治哲学等，将秩序建立任何普世的“共识”之上。这里就随之有两重导出的困难：一个是更高层级、更具包容性的“共识”是什么样的？另一个就是加入“没有共识，多元价值的人类是否能和谐相处？”这两个问题的深入解析，依然是当今复杂多变世界的核心问题之一。这里仅仅聚焦于这样一个问题：未来基于新共识的社会形态，将有哪些与众不同的特

---

<sup>31</sup> “确定性的寻求”，【美】杜威，《杜威全集》第四卷，华东师大出版社 2015，p.3

征？

控制论创始人维纳在 1950 年出版的一个小册子《人有人的用处：控制论与社会》中，转述英国 19 世纪诗人赫尔墨斯（Oliver Wendell Holmes, 1809-1894）的一首长诗“神奇的单马车（The Wonderful One-Hoss Shay）”中的观点：真正好的马车，是这个马车坏掉的时候，所有的零件都同时坏掉。一个轮子使用寿命将近时，车轮轴也不行了。这就是所谓的“恰当设计”——并不是说每一个环节都要设计得很棒，而是所有设计的匹配度极佳。<sup>32</sup>

区块链让这个社会不再追求 GDP、追求单边经济增长，而是追求“刚刚好”，也就是达到某种程度时就能说“够了”。在这样一个知道“足够”的社会，人的创造力才能得到无穷的释放，才能进入到艺术的、创新的氛围。在瑞典语中，有一个词语叫 Lagom，中文翻译为“有度”，据说这一词语来自维京时代。众人围坐在桌边，桌子上有一大杯啤酒，足够每个人喝一点。Laget om 这个词语就诞生了。意思是“轮流来”。演变至今，就成为 Lagom，并成为瑞典人日常生活中的口头禅。<sup>33</sup>

从这个意义说，区块链让每一个人达成自己的甜蜜三角——所能、所愿和所为之间实现良好匹配，不再有所愿非所

---

<sup>32</sup> 《人有人的用处：控制论与社会》，【美】维纳，商务印书馆 1978，p.45

<sup>33</sup> 《有度：瑞典人为什么自在》，【法】安妮·杜米佑，中信出版集团 2019，p.12

为、所能非所愿或者所做非所能的遗憾，而是尽快地进入到生命和生命共生演化的巨大网络之中，坦然地接受生命能量的相互支撑、流动，坦然地用自己的生命意愿去接触所有的生命意愿。

这听上去很浪漫，也很乌托邦。但回顾区块链的技术思想史，可以发现一个有趣的现象：区块链的奠基性技术（如时间戳、非对称加密、Hash 函数、默克尔树、共识算法、分布式系统等），贯穿着某种自控制论、信息论诞生以来长达 70 年的思想主线，这一主线从战后重建、冷战，到 60 年代的披头士文化、反战思潮，以及后现代思潮、硅谷崛起、微电子革命、太空探索、第三次浪潮、新经济，还包括黑客运动、好莱坞科幻大片、阿帕网、赛博空间、赛博格、赛博朋克运动等等，构筑出某种特色鲜明的历史画面。

这幅画面，从德国学者托马斯·瑞德的《机器崛起：遗失的控制论历史》一书中可见一斑。在《机器崛起》这本书中提到的三则“宣言”，就是极具代表性的文本。

第一个宣言，是 1985 年，加州大学圣克鲁兹大学哲学教授唐娜·J·哈拉韦（Donna Haraway, 1944- ），在《社会主义评论》上发表了《赛博格宣言（A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century）》。她认为，“一个赛博格就是一个

生控系统，是机器和有机体的一种组合，是社会现实，同时也是小说里虚构的生物。”（原书 1<sup>33</sup>）她用 C3I 这个著名的符号（指挥、控制、通信+情报），说现代战争是赛博格的狂欢，“我们都是赛博格”。

众所周知，赛博格<sup>34</sup>这个名词，是纽约洛克兰德州立医院的医生纳丹·克莱恩（Nathan Kline, 1916- 1983），和他的助手 Manfred Clynes（1925- ）在 1960 年的一篇文章《药物、太空和控制论：赛博格的进化》中杜撰的一个名词，这一名词被用作 D. S. Halacy（1919- 2002）1965 年出版的一本书的书名《赛博格》。

赛博格这个术语，可以看作维纳的控制论“失控”的一个标志，也可以看作控制论思想超越工程技术，进入有机体生命世界，进入政治学、社会学、人类学领域的一个分水岭。在这一概念的催化之下，60 年代涌现了大量新奇装备，包括即时感知设备、外骨骼系统、人机合体、增强触觉、头盔显示器、行走机器人等等。这一概念在当时的语境下，被认为是尼采“超人”理念的觉醒，是机器的觉醒，是这样一种信念：人类进步不再被动由生物进化所驱动，而是掌握在自己手中。

---

<sup>34</sup> 任何一个外部义肢装置都可以将一个人变成赛博格，比如作为技术社会化的载体的手机，它就是一个低层次意义上的赛博格，因为它利用网络平台（信息交换和连接）使得参与者可以与另外的参与者沟通和交流。资料来源：[http://cyborganthropology.com/What\\_is\\_a\\_Cyborg%3F](http://cyborganthropology.com/What_is_a_Cyborg%3F), 2010-12-24

1965年，英国数学家 Irving Jack Good（1916年12月-2009年4月）认为，“人类的存亡取决于超级智能机的早起建设”，1970年，科幻小说家弗诺·文奇（Vernor Vinge，1944年10月-）将“智能大爆炸”命名为“奇点”<sup>35</sup>。

2005年，美国技术预言家 Raymond Kurzweil 出版了《奇点临近》一书。他更在2009年成立了奇点大学，认为2045年“生物学意义上的人类将不复存在。”

可以说，赛博格这一概念，是超越传统机械还原论的两分法思想的一个重要标志。一方面，对生命的理解不再局限在有机体，整个自然环境都是活的，是充满生机的。1968年英国独立科学家、环保主义者詹姆斯·洛夫洛克（James Lovelock，1919年7月-）和美国生物学家琳·马古利斯（Lynn Margulis，1938年3月-2011年11月）提出的盖娅理论，充分体现了这一点。

第二个宣言，是1988年，蒂莫西·梅（Timothy May，1962-）发表的《加密无政府主义宣言（The Crypto Anarchist Manifesto）》。蒂莫西·梅的思想迄今仍可视为区块链“原教旨主义者”的精神教父。他认为技术将淘汰暴力、将拆除知识资产的栅栏、将打垮税收征管等等。伴随科技乌

---

<sup>35</sup> Singularity，大爆炸宇宙论所追溯的宇宙演化的起点。它具有一系列奇异的性质，无限大的物质密度，无限大的压力，无限弯曲的时空等。不少学者证明在广义相对论的宇宙学中，“奇点”是不可避免的，均匀各向同性的宇宙是从“奇点”开始膨胀的。资料来源：金炳华 主编.马克思主义哲学大辞典.上海：上海辞书出版社.2003:466

托邦主义<sup>36</sup>和反乌托邦主义的角力，出现了所谓“黑网”，认为宁愿将权力托付给数字、主张完全匿名、无物理地址访问，为确保数字投票、数字现金、数字自由的权益。

这一背景不但与当时的微电脑技术联手，而且呼应 1988 年计算机病毒大爆发、1991 年伊拉克战争的“沙漠风暴”行动，以及 1993-1994 年加密芯片的争议，1996 年 10 月跨国界网络入侵的“月光迷案”，出现了大量激进的文化思潮，包括 1996 年，约翰·巴洛(John Barlow, 1947 年 10 月- 2018 年 2 月)的《赛博空间独立宣言 (A Declaration of the Independence of Cyberspace)》。

《赛博空间独立宣言》，是该书提到的第三个宣言。美国《连线 (WIRED)》杂 2016 年 2 月的一则评论指出，电子前哨基金会 (EFF) 的联合创始人 John Perry Barlow，在 1996 年 2 月 8 日发表的这份《赛博空间独立宣言》，传递了一个清晰的信息：传统的政府，再也无法统治互联网。

Barlow 在宣言中告诉各国政府：政府在网络空间是不受欢迎的，政府对于网络空间没有主权。“我宣告，我们创建的全球社会空间，自然地不受你们强加给我们的独裁的束缚。你们没有任何道德权利统治我们，你们也没有任何强制办法，让我们真的有理由惧怕。”

---

<sup>36</sup> 乌托邦主义即空想社会主义。因十六世纪英国早期空想社会主义者托马斯·莫尔的著作《乌托邦》而得名。资料来源：李水海 主编.世界伦理道德辞典.西安：陕西人民出版社.1990:198-199

《连线》杂志的评论指出，在 20 年后的今天，我们却生活在一个日益“局域化”的互联网中，这包括美国国家安全局对美国本土，以及全球范围内进行的大规模监视、美国 FBI 去“暗网”执法的时代。Barlow 的独立宣言，也被一些人视为早年第一波互联网泡沫时，狂妄自大的产物；但报道中的 Barlow 本人，仍然对 20 年前的独立宣言深信不疑。不管怎么说，当今天的大数据、云计算、人工智能、虚拟现实、区块链、物联网、脑机接口等等技术又迎来强劲爆发的年代，Barlow 宣言中所触及到的问题，其意义和价值并未消失。

了解这一技术思想的演变史，有助于理解区块链技术背后的思想意蕴。表面上看，黑客精神、极客精神崇尚代码、崇尚机器的根由，与其说是探索新世界、新物种的冲动，毋宁说是对旧世界、旧物种的极度失望。科技乌托邦主义，宁愿将世界的主宰权交给机器，也不愿意交给政府或者宗教组织。在这个大的背景之下，区块链思想的乌托邦色彩，就昭然若揭了。

区块链试图编织这样一种庞大的“记录”网络，将人类的经济活动和社会活动都记录在案（技术的可行性暂且不论）。一旦记录在案，人的生存空间就由实体空间分化成多维的赛博空间。在这一虚实交织的世界里，整个世界的复杂性急剧爆炸。过去的知识谱系无法支撑未来世界的运转。那么区块链的伟大意义，就在于它率先提出了“测量”的问题，

即如何测量这个世界中所发生的一切“交易”？

互联世界的挑战用技术的语言说，就是“并发”。我们用一个通俗的例子说明这一点。

比如啤酒的生产与消费。从传统商业链条来看，一瓶啤酒的生产与消费，是一个“串行”的过程。从物流角度看，啤酒从原材料采买到流水线罐装，再到批发零售，抵达终端消费者；从资金链角度看，消费者扫码支付的10块钱，到了零售店小老板的口袋里，一个月之后，需要给上游批发商货款，批发商三个月之后给总代货款，总代半年之后给生产厂货款，这样一步一步的归集、交付。物流和资金流是逆向流动的，并且资金流永远“滞后于物流”，这一特点可称之为“串行”模式。

这一简化的模型，所反映出来的是这样一件事，即工业时代的财富分配是滞后于财富生产的。经济学家和政治家们为此构建了财富分配的三阶段理论：财富的第一次分配是工资，第二次分配是税收，第三次分配是社会救助与公益慈善。这是理解区块链带来生产方式、分配方式变革的一个现实基础。在工业资本主义时代，生产和分配首先是不同步的，财富的分配是滞后于生产，更重要的是财富的分配跟生产不但不同步，而且不对等。因此，马克思《资本论》分析资本主义生产关系时抽取出一个最重要的内核，就是“剩余价值”

和“劳动的异化”。

每一位参与劳动的劳动者的财富分配过程，受制于这样一个滞后、错位的生产、分配网络，受制于这样一种串行的信息连接、传递方式，大大掩盖了生产与分配的真实过程。这个问题在区块链之后，就有了一个非常巧妙的解决方法，具备了重新看待生产-分配网络的技术基础。

还是拿啤酒消费来举例。当消费者、零售商、批发商、生产商，所有供应链都处于同一个区块链网络的时候，这瓶啤酒是怎么消费的呢？当你一扫码这十块钱，并不是把钱给了小老板的口袋，这十块钱立刻分解成100个支付项：瓶子盖两毛钱，瓶子五毛钱，瓶子上刷的商标三毛八，瓶子里灌的饮料两块五，瓶子上分摊子的运输成本三毛二，瓶子上分摊的资金成本、银行利息两毛八……这个瓶子上凝结的物化劳动和物料成本，瞬间被分解成一百个碎片化的支付项，并且瞬间可以抵达财富分配的终点。这才是区块链的魅力所在。

传统商业社会的理论基础，是建立在信息不对称、存在交易成本的假设的基础之上的。并且这些理论因为缺乏相应的技术手段，无法从技术上建构一个“并发”的世界图景，只能将世界运转的基础，奠基在制度建设、治理体系、产业政策、货币政策等等的基础上。未来的世界是建立在“万物互联”假设的基础上的，区块链为这个万物互联的世界，又

提供了一个强有力的支撑：将经济行为、社会交往行为，奠基在一个孪生、并发的巨大网络中。一切并发的交易，理论上都可能被记录、被归集、被计算、被连接起来，成为物理世界无所不在的“数字孪生世界”。这幅画面的意义是深远的。

这个意义就在于：传统的生产、分配网络，是以“先生产、后消费、再分配”的串行逻辑运转的，而奠基于区块链网络的新生产、分配网络，则是以“边生产、边消费、边分配”的并发逻辑运转的。这是一场史诗般的伟大变革。区块链成为构建新的生产关系的重要基础设施。

这样并发的生产和分配图景出现后，如何导致一个“知足社会”图景呢？这里有一个资源消耗总量的概念。从人的这个基本需求来说，也就是人一生中消耗掉的资源总量来说，其实是一定的。有人大概估算过，我们今天的生产已经远远超过生存所需，只不过我们过去的社会运转逻辑，经济学的基本假设，是建立在“稀缺”、“自私”等概念之上的。这样的假设在万物互联的未来图景下，需要有全新的思考。即便是从乌托邦的角度看，这样的思考也富有启发意义。

研究贫困问题的经济学家阿马蒂亚·森认为，对于每日生活所需的人而言，今天的矛盾焦点并不是无法生产出足够的生活必需品，而是这些已经生产出来的生活资料的分配问

题<sup>37</sup>。建立在稀缺假设和自利的经济人假设基础上的经济学，再加上轴心时代“金规则”的价值观，使得今日世界的基本逻辑，依然是“零和博弈”的逻辑，依然是一个占有的社会。

在区块链以前，第一次、第二次、第三次工业革命都在提高生产效率，可是到了第四次工业革命，问题的焦点发生了变化，生产不再是舞台中央，公平问题日益成为新政治经济学的核心问题。所以区块链的经济思想和社会价值，是否可以总结为这么几句话：马克思曾经预言过的“自由人的自由联合”，已经通过互联网真实地出现在地平线上；人的劳动和财富创造之间的物化关系，日益解耦；创造和创新成为人的日常生活的必要组成部分；生产活动和价值分配的权利，第一次有了完整统一的可能，这个统一的基础，就是区块链。

---

<sup>37</sup> 【印】阿马蒂亚·森，《贫困与饥荒》，商务印书馆，2001-1

# 人工智能

# 清华——工程院等联合发布 《人工智能发展报告 2020》

2021年4月由中国工程院-清华大学知识智能联合研究中心、清华大学人工智能研究院和中国人工智能学会联合发布《中国人工智能发展报告 2020》，其核心发现主要有以下10点：

- 1、总结了过去十年的十大人工智能研究热点；
- 2、过去十年有五位人工智能领域学者获图灵奖；
- 3、顶刊获得最多奖项是计算理论、安全与隐私和机器学习；
- 4、人工智能最高引文的头部主题为机器学习、计算机视觉领域；
- 5、中国在多个相关领域紧跟美国，处于世界前列，在多媒体和物联网领域超过美国，部分领域还需追赶；
- 6、全球范围内中国人工智能学者数量占 9.8%，美国是 62.2%；
- 7、清华大学是唯一入选全球人工智能高层次学者数量 Top 10 的中国机构；
- 8、京津冀、长三角、珠三角地区为国内主要高层次人工智能人才聚集地；
- 9、过去十年中国专利申请量是世界第一，是第二名美

国的 8.2 倍；

- 10、人工智能未来重点发展技术方向包括：强化学习、神经形态硬件、知识图谱，智能机器人、可解释性人工智能、数学伦理和知识指导自然语言处理等。

# 清华——工程院等 发布人工智能未来重点技术方面的商榷

中国开源软件推进联盟（COPU）

2021. 1. 15

看到由清华与工程院专家撰写的《人工智能发展报告2020》，深受启发，但在读到人工智能未来重点发展技术方面，似有商榷之处：

1、国内人工智能研发通常忽略机器学习/深度学习处于不可解释状态，认为其潜力已经挖尽！

其实正是它支持了今天人工智能的繁荣。上世纪末抗生素产生了抗药性，麻省理工的科学家通过深度学习模型，对2300多种抗生素药物进行训练，研发出能够解除抗药性、杀死耐药的超级细菌的全新抗生素；谷歌的科学家利用深度学习训练蛋白质，研发用于基因医疗的 AlphaFold；英国基于机器学习研发了全球首起由人工智能控制的喷气式高超音速六代机——“暴风雨”战斗机（原型）；国内外一批科学家基于机器学习筛选，用于研发新材料，新药物；自动驾驶、无人驾驶发展很快，它也是基于机器学习发展起来的。

2、谈到人工智能未来技术的发展，重点报告中虽然也提到了可解释性人工智能，但未突出！联盟陆首群主席在文章

中做了补充，两位清华大学教授强调的意见：2020年12月，沈向洋教授提出，我们需要拥抱开源，我们的创新第一是人工智能，最重要的事情要做可解释的人工智能。2021年1月，姚期智院士提出：人工智能第二大技术瓶颈是机器学习算法缺乏可解释性，很多算法处于黑盒子的状态，亟待突破。

3、报告中谈到未来发展重点：神经形态硬件、知识图谱、知识指导自然语言处理，是否可调整为：神经形态硬件，异步脉冲神经网络及神经形态计算系统。知识图谱、知识指导自然语言处理，是否可改写为：大规模语义网络（知识图谱），而且其支持实现认知智能还差最后一公里。

报告中赞扬：清华大学是唯一入选全球人工智能高层次学者数量 Top 10 的中国机构。对此国人是持有异议的，清华大学、北京大学培养的本科生大部分赴美未归，在硅谷的清华出身的 IT 专家（大多搞人工智能研发）约两万人，不过2016年以后出现了人员回国潮。一些回清华的资深海归高层次学者如姚期智、沈向洋、沈寓实、王卓然等，是否统计在报告中所列的 Top 10 之内？

# 自动驾驶和无人驾驶

## 编者的话：国内外自动驾驶和无人驾驶<sup>38</sup>

目前，国内外自动驾驶和无人驾驶发展很快，它们是基于机器学习（无解释性）发展起来的，竞争激烈。

我们先介绍国际自动机工程师学会 2004 年制定的“无人驾驶和自动驾驶技术界定标准”，将路况分成 5 级：

- 0: No Automation 人工驾驶
- 1: Driver Assistance 辅助驾驶
- 2: Partial Automation 半自动驾驶
- 3: Conditional Automation 高度自动驾驶
- 4: High Automation 超高自动驾驶（人工接手）
- 5: Full Automation 全自动驾驶（完全无人）

国内外自动驾驶和无人驾驶主要是两个方案：一是单车智能方案，通过在车上安装多个雷达传感器，实现对路况环境感知，采用自动驾驶行为向量预测算法模型

(VectorNet)，面对周围车辆和行人的行为不确定性及各种规则之外的行为（特别在遇到盲点与遮挡时），在行进中，在未来 5 秒内提高行为预测；二是车路协同方案，在路边配置、安装激光雷达传感器，由车辆与路边交互而完成对周围环境的感知（这个方案人称“中国方案”，百度、长沙市、一汽三家合作，正在测试车路协同方案）。

---

<sup>38</sup> COPU 编者陈伟、鞠东颖。

自动驾驶和无人驾驶现正在形成相当规模的行业，在美国，围绕 Waymo、Zoox、Cruise、ArgoAI、Aurora 硅谷五大家属由几十家企业组成新行业；在中国，百度仍是领先的！百度拥抱开源，在发布 Apollo 6.0 版本时，来自 97 个国家 5.5 万名志愿开发者协同开发、修改自动驾驶软件 150 万行，并发展生态合作伙伴 210 家企业和科研机构。

在中国进行自动驾驶和无人驾驶研发的还有腾讯、蔚来、四维图新、小马智行、北京三快、领骏科技、华为、滴滴出行及若干传统汽车厂商，其中一些企业也积极参加国际赛事，并在单项指标或综合指标上创优。国内新行业也发展到十几家。

AR/VR/XR

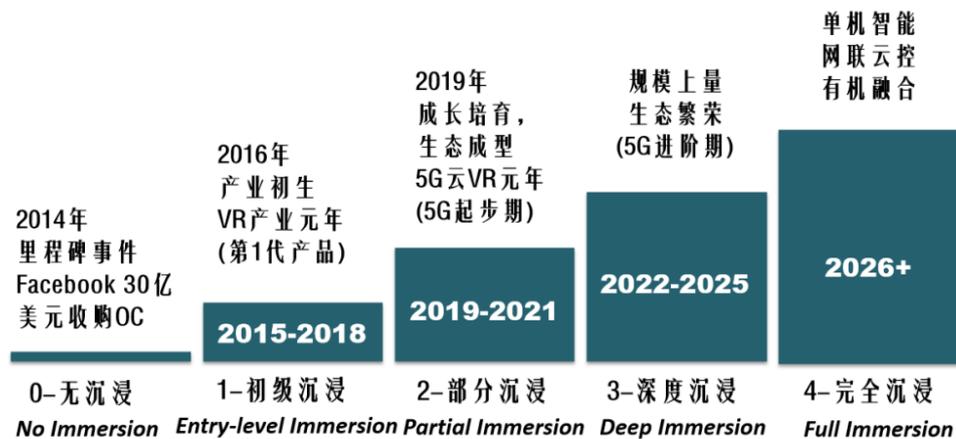
# VR 创新与实践

中科创达 刘寿永 邹鹏程

## 一、VR 当前的产业现状

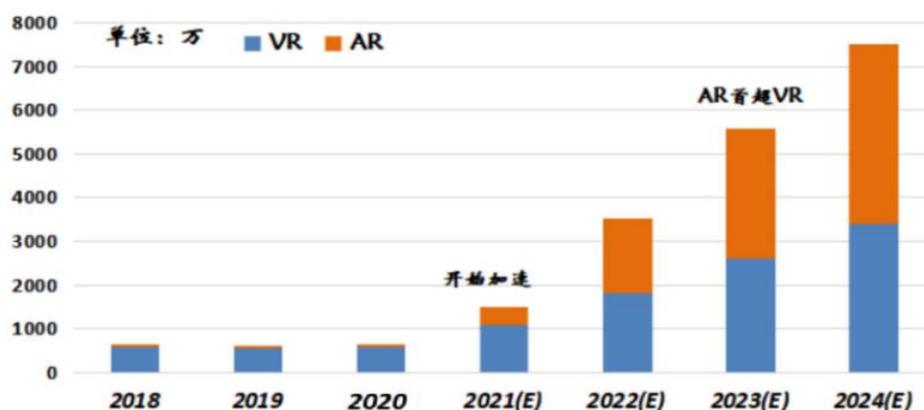
从广义来看，虚拟现实包含增强现实（Augmented Reality, AR），狭义而言彼此独立，本文统称为 VR。虚拟现实旨在使用户获得身临其境的沉浸体验，依据中国信通院《虚拟（增强）现实白皮书（2018）》中分级标准，可将虚拟现实划分为五个发展阶段，不同发展阶段对应相应体验层次，继 2016 年虚拟现实产业元年、2018 年云 VR 产业元年、2019 年 5G 云 VR 产业元年过后，2020-2021 年将成为虚拟现实驶入产业发展快车道的关键发力时窗，目前全球处于部分沉浸/成长培育期。其中，在终端设备方面，开始规模上量，适配场景与功能定位体系日益清晰完备，例如华为 VR Glass、Focal 等轻量级 VR/AR 终端通过强化通信连接能力，以及摄像头提供虚拟助手等功能进而变身为手机伴侣，微软 Hololens2 等高性能一体式 AR 终端可在一定程度上取代 PC，作为新兴生产力平台，Facebook Quest2 等高性能 VR 终端可作为电视与游戏机等传统文娱平台的产品演进形态。在内容应用方面，题材形式日益丰富，内容与特定终端平台加速解耦，内容开发、调试与营销工具渐趋成熟，可自给、能盈利的内容生态开始成型，例如，标杆企业 Facebook Quest 平台

内容收入已达到 1.5 亿美元，35 款游戏收入达到百万美元，沉浸声、手势识别与虚拟化身等特色内容制作软件开发工具包（Software Development Kit, SDK）陆续发布。在网络平台方面，2020 年成为 5G 创新业务从 0 到 1 实现跨越的关键窗口，作为 5G 时代首要的创新业务，一方面，VR 为 5G 这一国家新型基础设施提供了普适典型的应用场景，另一方面 5G 有望打破单机版 VR 小众化的产业发展瓶颈。2019 年期间，北京、上海、广州、青岛、杭州、成都、福州、南昌、沈阳等地方政府已将 5G 云 VR 提上工作日程，相继编制或正在实施专项政策与相关工程。



目前，虚拟现实终端出货量稳步增长，AR 与一体式增速显著，不同终端形态间的融通性增强。受新冠疫情及宏观经济形势影响，据互联网数据集团（International Data Corporation, IDC）统计，2020 年全球虚拟现实终端出货量约为 630 万台，VR、AR 终端出货量占比分别 90%、10%，预计 2024 年终端出货量超 7500 万台，其中 AR 占比升至 55%，

2020-2024 年五年期间虚拟现实出货量增速约为 86%，其中 VR、AR 增速分别为 56%、188%，预计 2023 年 AR 终端出货量有望超越 VR。比之 2018-2020 年相对平缓的终端出货量，随着 FacebookQuest2、微软 Hololens2 等标杆 VR/AR 终端迭代发售以及电信运营商虚拟现实终端的发展推广，2021 年有望成为虚拟现实终端规模上量、显著增长的关键年份，VR/AR 终端平均售价将从当前 2500/9700 元人民币进一步下降。此外，华为 VRGlass 手机伴侣、PicoNeo2 等一体式头显终端均可通过串流功能而不再受制于移动平台的功耗预算与渲染算力，跨终端形态的使用融通性显著提高，一体式终端出货量份额预计将从 2020 年 51%进一步升至 2024 年 64%。



其次，我国积极推动虚拟现实产业发展。自 2016 年虚拟现实被列入“十三五”信息化规划、互联网+等多项国家政策文件以来，国务院、国家发改委、工信部、科技部、教育部等部委相继出台指导政策支持虚拟现实产业发展。国务院从十三五规划开始把虚拟现实视为构建现代信息技

术和产业生态体系的重要新兴产业，在《新一代人工智能发展规划》中将虚拟现实智能建模技术列入“新一代人工智能关键共性技术体系”，2020年相继出台《关于进一步激发文化和旅游消费潜力的意见》、《新时代爱国主义教育实施纲要》和《关于推进对外贸易创新发展的实施意见》等文件，进一步明确虚拟现实在文化旅游、教育宣传、商贸会展等领域的创新应用。在2021年出台的《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》中将VR/AR产业列为未来五年数字经济重点产业之一。

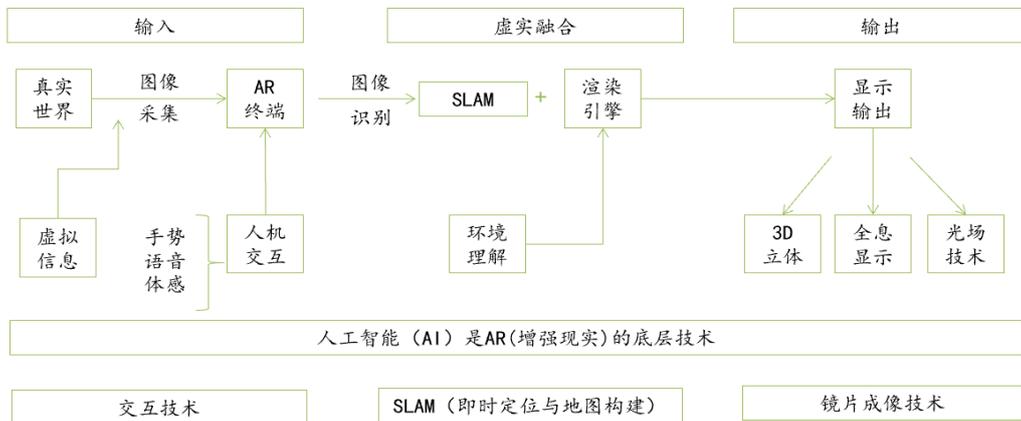
## 二、VR的核心技术

VR融合了真实世界和虚拟世界的信息，反馈给观察者，同时可以做到实时交互，这就形成了虚拟现实跨界复合的技术特性，VR本身具有几个特性：信息叠加，空间定位，实时交互。这几个特性决定VR的主要技术是指近眼显示、感知交互、网络传输、渲染计算与内容制作，通过梳理五大领域下辖的各细分关键技术，可归纳总结如下虚拟现实技术体系，其中各技术点的发展成熟度具体如下：近眼显示方面，快速响应液晶屏、折反式(Birdbath)已规模量产，Micro-LED与衍射光波导成为重点探索方向。渲染计算方面，云渲染、人工智能与注视点渲染等技术进一步优化渲染质量与效率间

的平衡。内容制作方面，WebXR、OS、OpenXR 等支撑工具稳健发展，六自由度视频摄制技术、虚拟化身技术等前瞻方向进一步提升虚拟现实体验的社交性、沉浸感与个性化。感知交互方面，内向外追踪技术已全面成熟，手势追踪、眼动追踪、沉浸声场等技术使能自然化、情景化与智能化的技术发展方向。网络传输方面，5G+F5G 构筑虚拟现实双千兆网络基础设施支撑，传输网络不断地探索传输推流、编解码、最低时延路径、高带宽低时延、虚拟现实业务 AI 识别等新兴技术路径。而所有技术当中，最为重要的是显示技术，下图以 AR 为例，来展示 AR 的显示技术原理：

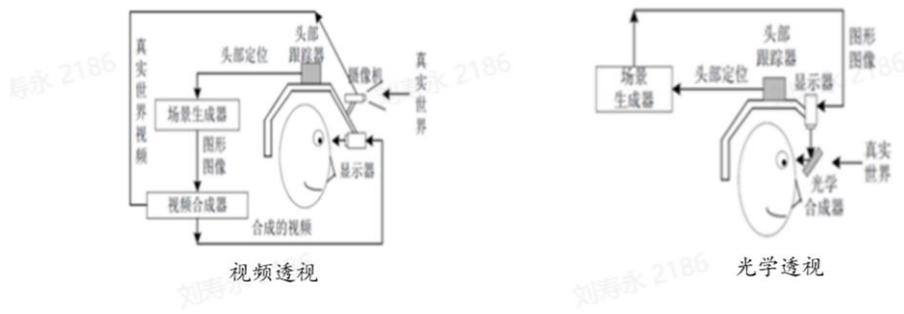
## 技术原理

AR系统包括数据处理、SLAM、显示、人机交互等技术环节，通过摄像头获取真实环境信息，结合传感器进行定位跟踪、交互，通过显示设备生成虚拟场景，叠加到现实场景。



显示技术，主要是由以下的几个技术组成：第一，透视方式，可以分为视频透视和光学透视，下图是两种显示方式的原理和特点：

AR眼镜按真实环境的表现方式可以分为，视频透视式 (video see-through) 和光学透视式 (Optical see-through)。



- **视频透视**：类似于VR加上摄像头，易于实现，说到底它并不依赖眼睛真实地去看外界的事物，它只是一台有着实时外部监控画面显示的 VR 设备。有时候被引入新概念称作 MR (Mixed Reality)，形式大于意义。
- **光学透视**：跟实际场景结合的更紧密，真实感更强，目前大部分厂商采用的光学透视式方案。

第二，近眼显示模块，主要有基于硅基液晶，有机发光二极管，数字处理等 LCD 技术，由不同的厂家提供，不同的模块有不同的优缺点，比如 LCOS 结构简单，但厂家多，标准不统一；LCD 亮度低，但生产工艺成熟。根据不同的应用场景来选择不同的近眼显示模块。下表展示了不同显示模块的区别：

	LCOS 硅基液晶	LCD 液晶显示技术	OLED 有机发光二极管	DLP 数字光处理
组成	硅基液晶 + 棱镜组，是一种基于反射模式，尺寸非常小的矩阵液晶显示装置。谷歌眼镜也用到了这种技术，特点是在 IC 芯片上集成超精细度的像素点，精细度可达3000dpi。	在两片平行的玻璃当中放置液态的晶体，两片玻璃中间有许多垂直和水平的细小电线，透过通电与否来控制杆状水晶分子改变方向，将光线折射出来产生画面。	结构：薄而透明具半导体特性之铟锡氧化物 (ITO)，与电力之正极相连，再加上另一个金属阴极。	DMD芯片 + 旋转滤色片，这种技术要先把影像信号经过数字处理，然后再把光投影出来。DLP的数字性质可以获得具有精确灰度等级的图像质量以及颜色再现。
厂商	Himax OV SONY EPSON	天马	Sony /BOE	TI
理论特点	结构简单，可采用CMOS工艺大量生产，成本低，亮度高，像素密度最高，功耗低。	亮度低，(同等亮度下) 功耗高，像素密度低，反应速度慢。	色彩鲜艳，亮度高，功耗低	亮度高，像素清晰(相对LCD) 反应速度快，工艺复杂，色彩不如LCD鲜艳。
现实情况	厂家多，标准不统一，生产工艺不成熟，产量小，价格高，低功耗的优点未能发挥。	由于生产工艺成熟，价格低，功耗低，延迟缺陷并不明显，像素密度较难提高。	价格相对较高，像素高。	主要用于投影仪、微型投影仪，功耗较大。

第三，是光学系统，光学系统的的质量决定了成像质

量，下面是四种常用的光学系统，立轴光学，棱镜，曲面棱镜和波导，下图是针对不同光学系统的一个对比：

	1	2	3	4
	离轴光学	棱镜	曲面棱镜	波导
厚度	头盔式	> 10mm	> 8mm	超薄
视场角	-	15°	30°	30°~60°
技术壁垒	很低	中等	较高	极高
公司	Meta	Google Glass 影创	亮风台 蓝斯特, Epson	灵犀 Magic Leap, Lumus
				

不同的显示模块和光学系统，走不同的技术路线，应用到了不同的领域，下表是一个对近眼显示和光学系统总结：

品牌	技术路线	主要市场	品牌	技术路线	主要市场
Meta	离轴反射	综合	Oglass	离轴反射	工业
ODG	棱镜光学	企业、游戏	亮风台	自由曲面光波导	工业、游戏
Google	棱镜光学	制造业、物流	蓝斯特	自由曲面棱镜	工业、游戏
EPSON	自由曲面棱镜	医疗、工业	影创科技	自由曲面棱镜	工业、教育
DAQRI	阵列波导	工业	悉见科技	自由曲面棱镜	文化旅游
Atheer	阵列波导	综合	泉龙科技	自由曲面棱镜	智能安防
Lumus	阵列波导	综合	耐德佳	自由曲面光波导	企业版
HoloLens	全息波导	综合	灵犀微光	阵列波导	工业
Sony	全息波导	工业	珑璟光电	阵列波导	工业

### 三、中科创达在 VR 上的实践

中科创达针对虚拟现实领域推出的（VR）SoM 核心计算模块，配合专为 VR 产品开发的 VROS 以及一系列创新技术，为全球多家知名客户提供了软硬一体的解决方案，已

服务包括 Facebook、谷歌、索尼、松下、Rokid、联想、Vuzix、亮风台、微鲸等知名厂商。中科创达能够在 VR 相关领域提供 6 项服务。分别为提供高通平台的硬件方案，包括定制开发以及公板方案并交付主板；提供 VR 整机交付服务，包括头戴式设备的结构设计、散热设计以及天线设计，生产组装以及整机测试；提供平台软件开发，除底层 BSP（Board Support Package，板级支持包）调试以及优化外，还可以提供针对虚拟现实设备的 Android OS 定制优化，包括系统裁剪、VRSDK、性能以及功耗的优化；提供 VR 开发套件，用于前期原型验证；提供 VR 应用开发，包括 Launcher 开发以及行业应用开发；提供 Kanzi 工具及 VRUI 设计软件开发。

在 VR 市场上，目前国内机会比较少，主要原因在于，VR 市场设备形态比较固定，目前市面上有不少高品质低价格的 VR 产品，例如 Oculus、HTC 和 Pico 等，无论是 C 端还是 B 端，都会选择直接采用现有的成熟设备，再去开发应用或者内容，不太可能会单独开发硬件设备，从目前接触到的客户来看，韩国市场主要还是以 VR 为准，且都属于设备自研类型的，如果使用高通平台，我们还是有很大机会。

AR 市场以 ToB 为主，由于目前 AR 关键技术还不够成熟，以现有的技术解决一些行业需求，可能会存在不同形

态的 AR 设备，这样会带来大量的定制化需求。AR 的前景是大家一致看好的，那么一些传统企业也会在 AR 技术上不断探索，可能会带来一些验证性的 AR 项目机会。由于 Google 和 Apple 在手机上推出的 ARSDK，加速了基于摄像头的 AR 应用市场的发展，未来手机标配 DP 后，可以直接搭配 AR 眼镜，将 AR 应用从 2D 平面显示转化为 3D 立体显示，手机厂商可能会推自己的 AR 眼镜配件。与 AR 核心技术供应商合作，提供完整解决方案给 B 端客户，例如与 ARSDK 或者光学厂商合作，他们拥有相当丰富的 AR 客户资源。

从客户需求上来讲，我们主要先看下目前 VR 设备本身存在哪些问题：1. 重量重，2. 体积大，3. 发热大，4. 续航短，5. ID 设计难，6. 亮度低，7. 视场角小，8. 色彩还原差，9. 分辨率低。那如何来解决这些问题是我们必须面对，也是对我们提出的技术挑战，我们先看下目前 VR 设备的形态一些优缺点。

从机型来看，目前的主要形态是一体机和分体机，一体机的主要优点在于无线，便于携带，无需外接 PC，价格相对便宜，缺点在于移动处理器的性能相对不足，图形渲染能力不足；而分体机的主要优点是性能好，6DoF 稳定，而缺点是价格贵，需要外界 PC，通过两者的对比一体机价格比较适中，性能偏弱，一般会用于观影，教育，培训等领域，而分体机能够提供更好的 VR 体验，主要面向游戏爱

好者，市场基本被 HTC，OCULUS，Sony 和 Microsoft 瓜分。



**一体机**



**分体机**

一体机从光学系统来分，有双面自由曲面和单目自由曲面，双目自由曲面主要优点携带方便，支持 VSLAM 定位，缺点是重量大，发热大，续航短，光学体积大，FOV 小等；而单目自由曲面主要是小巧轻盈，佩戴舒适，但也有不足，单眼显示，不支持 VSLAM，和双目自由曲面同样有光学体积大，FOV 小的缺点。而分体机主要应用了双目光波导的光学系统，具有运算能力强，续航时间长，体积小，发热轻，镜片薄等优点，但同样有 FOV 小等不足。



**双目自由曲面**



**单目自由曲面**



**双目光波导**

为了解决上述的问题，增加更好的用户体验，创达通过合作和自研的方式来攻克核心技术，器件上还要从显示屏，光学系统等方面选择合适的供应商，算法通过和合作伙伴一起对 6DoF 手柄，手势识别，语音识别，眼球追踪等进行深度优化，同时在基础 OS 层面通过 2D 分屏，畸变矫正，功耗，实时调度等技术的引入，极大的提供整体的显示性能，同时通过对电池的优化，主要是结构设计和软件控制，达到增加续航能力的目的，最后提供简单易用的 SDK，方便开发者和客户做二次开发，下表是关键技术和合作伙伴描述：

类别	关键技术	描述	供应商/合作伙伴	说明
器件	显示屏	OLED/LCD/DLP/LCOS	BOE/Sony/	
	光学	离轴/棱镜/自由曲面/波导	耐德佳/珑璟/惠牛/灵犀	
算法	VIO/SLAM	单目/双目/TOF	高通/Occipital/Google	
	交互	3DoF/6DoF 手柄	Ximmense/HTC	
		手势识别	Crunchfish/Leapmotion	
		语音识别	科大讯飞/Rokid/思必驰	
	眼球追踪	Tobii/七星易维		
软件	SDK	VR/AR SDK	高通/视+/亮风台/Nibiru/Google/TS	
	系统优化	2D分屏/畸变/裁剪/性能/功耗/定制化开发/低延时	TS	
硬件	电池方案	分体式电池方案/双电池方案	TS	
	结构设计	散热	TS	
		人体工程设计	TS	
		配重设计	TS	

伴随 5G 技术的出现，有机会解决 VR/AR 在技术、产业链未突破前的痛点，但是目前时机不成熟，原因如下：针对游戏场景，1) 5G 传输依然有延时，存在额外带来的 4ms 的延时，传输协议需要优化；2) 新建片区边缘计算的高性能机房，针对观影场景，主要是单项传输，延时不会影响

使用，但是缺少优质的 VR 视频内容，目前没有视频或者比较少有的视频是非要拿 VR 看不可的，针对 AR 产品，主要落地的场景为工业领域的作业辅助，可以释放双手进行作业的同时，可以通过 AR 眼镜获取信息以及帮助，这个场景对设备性能没太大要求。

通过对操作系统深度优化和定制来构建时候 VR 的软件平台。首先，从 BSP 底层做深度优化，把 CPU，GPU，NPU 和网络以及内存等硬件的性能榨干；其次，对 Android 系统进行深度裁剪，剔除其中无用的模块，使系统的启动时间控制在 10s 以内；第三，优化和 VR 相关的核心功能，通过减少缓冲复制，图形渲染算法等来加速显示时间；第四，矫正镜头变形，第五，通过 IMU 和摄像头的数据融合，对物体进行追踪和定位。最后，通过提供 VR 的 SDK 来帮助开发者和企业，开发 VR 应用程序。

## 未来发展

伴随着技术的日趋成熟，产业的不断壮大，生态的逐步完善，内容的逐渐丰富，未来的 VR 必定在新基建和数字化转型的大潮中发挥更大的作用，期待更多的产业链上的合作伙伴同心协力，在时代的大潮中，在国家政策的引领下，发展围绕 VR 的制造业，打造技术高地，推广研究成果，培养人才等方面占领战略制高点，并能广泛应用到商贸，工业生产，医疗健康，教育培训，文化娱乐等多个行

业，对公共服务，拉动就业等方面会起到显著效果。

# 开源教学与实践

## MIT 开设的一门网络实验课程——操作系统工程

编者的话<sup>39</sup>：2008 年 COPU 曾向某些国内大学推荐 MIT 网站开设的操作系统实验型课程，这门课程的特点是：开源，具有教学、设计、实验（包括培养学生动手能力）全部内容，用时一个学期（半年），非常适合学生完成一个毕业设计（当然也可用作课程教育），适于计算机专业本科生。

大家可查阅 MIT 网站：

<https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-828-operating-system-engineering-fall-2012/>

我们邀请陈渝老师对最新版的 MIT 网站——操作系统工程作一个介绍（发现不少中美大学正在使用这个网站），供国内高等学校参考。

---

<sup>39</sup> COPU 编者陈伟、鞠东颖。

# Xv6 操作系统工程课程介绍

清华大学陈渝

## 一、概述

在 2002 年秋季，Frans Kaashoek, Josh Cates, and Emil Sit 在 MIT 开设了一门新的实验型课程“操作系统工程”，英文名称是“Operating Systems Engineering”，课程代号是“6.097”，后改为“6.828”，在此课程上，一开始采用了“莱昂氏 *UNIX* 源代码分析”（英文书名是“Lion's Commentary on UNIX 6<sup>th</sup> Edition With Source Code”）作为参考资料。此参考资料描述的 UNIXv6（简称 V6）是运行在古老的 PDP-11 计算机系统中。为了让学生更好地理解 V6 的实现，Frans Kaashoek 等从 2006 年夏季开始，参考 V6 的架构，在 x86 计算机系统中重新实现了一个支持多处理器计算机系统的类似 UNIX 的教学用操作系统，称为 Xv6。在目前的 MIT 本科生课程“6.828: Operating Systems Engineering”中，Xv6 主要用于讲课，而另一个基于 exokernel 架构的 JOS 主要用于做试验。目前 Xv6 在 MIT 的网址在 <http://pdos.csail.mit.edu/6.828/Xv6/>

与现代更先进的 Linux 或 BSD 操作系统不同，Xv6 操作系统内核简单到可以在一个学期内完成，但仍然包含 UNIX

中重要的重要概念和组织。而且 Xv6 操作系统内核可以自带文档。Xv6 的 Makefile 的一个功能是可以选择以可读格式生成整个源代码列表的 PDF 文档，且整个文档（包括交叉引用）只有 99 页。这让人想起最初的 UNIXV6 源代码，它以类似的形式发布在 Lions' Commentary on UNIX 6<sup>th</sup> Edition 中，并带有源代码。

Xv6 采用了类似 MIT 协议进行开源，已被用于许多大学的操作系统课程，包括清华大学，北京大学，哈工大深圳分校，南开大学，利兹大学、石溪大学、西北大学、乔治华盛顿大学、东北大学、耶鲁大学、哥伦比亚大学，Ben-Gurion 大学，约翰霍普金斯大学，波特兰州立大学，威斯康辛-麦迪逊大学，宾厄姆顿大学，犹他大学，佐治亚理工学院，伊利诺伊大学芝加哥分校，美国罗格斯大学等。

## 二、架构与组成

### 1、总体设计思路

操作系统是一种软件，操作系统没有一个精确和统一的定义。操作系统是一种比较复杂的软件，我们可以从多种角度来了解操作系统。从操作系统的任务来看，操作系统的任务主要是控制和管理计算机系统硬件资源并对应用软件和用户提供各种方便使用计算机的功能。通过操作系统，能有效地组织和管理计算机系统硬件资源和其他软件资源，向用户和应用软件提供各种服务功能，使

得用户和应用软件能够灵活、方便、有效地使用计算机，并使整个计算机系统能高效地运行。Xv6 采用了典型的 UNIX 单体内核的总体设计方式，内核运行在用户态。应用程序通过系统调用访问内核的服务。总体上，Xv6 包括系统调用接口、进程管理、内存管理、文件系统、I/O 管理等几个主要模块。

系统调用是应用程序访问操作系统的接口。在系统调用接口上，通用操作系统与基于此操作系统的应用程序处于两个不同的 CPU 特权态，操作系统处于核心态，而应用程序处于用户态。在核心态可以执行 CPU 特权指令，而用户态无法执行特权指令，且只能通过特定的指令或中断来访问操作系统提供的各种功能。这在一定程度上保证了系统整体的安全，避免应用程序对操作系统可能的破坏。

在内存管理方面，通用操作系统采用了虚拟内存管理方式，这样可以使内存需求超过实际物理内存的进程/线程能够执行，其主要思想是把重要和常用的数据和执行代码放在物理内存中，把不常用的数据和执行代码放到二级存储（这里主要指的是硬盘等可在掉电后保存数据的存储介质），随时根据系统执行情况替换放在内存中的数据和代码。而且通过虚存管理可以实现对不同内存区域的保护，不同进程之间，或者应用程序和操作系统之间的地址空间相对隔离。这样一般情况下不同进程的地址空间不能直接

访问，且应用程序不能直接访问内核地址空间。所以一个与错误的应用程序不会导致系统的崩溃，从而增加了系统的可靠性。Xv6 操作系统没有采用虚拟内存管理，而是采用了简单的基于 X86 段模式的单一地址空间管理方式。在内存分配和释放的管理上，Xv6 相对实现得比较简单，采用基于可变分区分配的首次适配算法，容易产生内存碎片。

在进程/线程管理方面，当前通用操作系统结合虚存管理，采用进程和线程结合的管理方式。进程代表了一个程序执行的过程以及其所占用的计算机资源（包括 CPU、内存、文件等），进程的执行流可用线程来表示。操作系统的调度单位可以是进程或线程。一个进程可以包含多个线程，属于同一进程的多个线程共享进程管理的资源，比如属于同一进程的多个线程共享进程所管理的内存，这样这些线程可以直接访问属于进程的全局地址空间。Xv6 操作系统实现了一个基于进程（没有实现线程）的简单进程管理机制。

在文件系统管理方面，当前通用操作系统结合虚存管理，实现了多种复杂、高效且可靠的文件系统，且建立了一个统一的虚拟文件系统层，屏蔽不同文件系统的差异，对上层提供统一的接口。且与用户管理和进程管理结合，可实现安全管理，保证对文件的安全访问。Xv6 操作系统实现了一个相对简单的基于 inode 索引方式的文件系统。

在 I/O 管理方面，Xv6 操作系统与通用操作系统（特别是类 UNIX 操作系统）差别不是特别大，都把设备“看成”是一种特殊的设备文件，有设备号，用文件的访问接口来进行打开、关闭、读、写和控制等操作。在灵活性方面，Xv6 驱动程序不能象通用操作系统那样根据硬件情况动态加载，而是在编译的时候就静态确定的。

## 2、总体架构

从操作系统模型上来看，Xv6 是一个单地址空间的层次式单体内核，不是微内核（micro kernel）模型的操作系统（如 Mach，QNX），与通用操作系统（如 Linux）的架构在地址空间和特权模式上也有一定的差别。下面主要分进程调度、内存管理、同步互斥、文件系统几方面对 Xv6 进行介绍。

### • 同步互斥

由于在 SMP 架构中，内存磁盘等硬件资源在所有 CPU 中都是共享的，所以在需要某种机制对资源进行互斥访问控制。在 Xv6 中，通过实现了 spinlock，从而可以对共享资源加锁来限制同时访问此资源的 CPU 数量。

### • 内存管理

在内存管理方面，Xv6 采用了段式虚拟内存的管理方式。每个用户进程所占用的内存都是在一个连续的段中。用户进程内存的分布为：代码段、静态变量段、固定大小

的栈和可变大小的堆空间。由于进程内存是按照段管理的，因此在每次分配进程内存时，Xv6 将找一片正好能放下整段的连续内存块进行放置。

- 进程管理

因为是基于 SMP 架构，操作系统中的多个进程会占用计算机系统多个 CPU 执行其具体功能，由于进程数量大于 CPU 数量，这就涉及到进程如何分时共享 CPU 的操作系统管理问题，具体包括如何创建进程、如何删除进程、选择哪个进程占用哪个 CPU，何时进行进程切换，进程能够持续占用 CPU 的时间片段的大小设定等。在 Xv6 中，首先其进程是基于时间片来调度的。每次进程的调度是由时钟中断产生的，或者是因当前进程主动放弃。其次，每个 CPU 之间都共享一个进程池（具体实现为一个全局数组），其中有所有待运行的进程。在每个时间片中，CPU 将当前运行的进程放回进程池，然后从进程池中选取另一个待运行的进程进行执行。

- 文件系统

Xv6 中提供了一个简单的文件系统，这个文件系统提供了大多数 POSIX 标准的接口。由于这个文件系统比较简单，其中一个文件最多由  $(12+128)$  个组成，所以文件的大小也被限制在  $(12+128)*512\text{Bytes}$ 。在这个文件系统中提供

了一个 Buf 层，用来缓存磁盘上的数据。但是此文件系统是写直达的，因此每次更新都会直接写到磁盘上。

### 三、实验设计方面

Xv6 操作系统的实验设计主要包括如下一些内容：

1) 启动操作系统的 bootloader，用于了解操作系统启动前的状态和要做的准备工作，了解运行操作系统的硬件支持，操作系统如何加载到内存中，理解两类中断——“外设中断”，“陷阱中断”等，以及系统调用等；

2) 物理内存管理子系统，用于理解 x86 分段/分页或 riscv-64 分页模式，了解操作系统如何管理物理内存；

3) 虚拟内存管理子系统，通过页表机制，以及中断——“故障中断”、缺页故障处理等，了解基于页的内存替换策略；

4) 内核线程子系统，用于了解如何创建相对与用户进程更加简单的内核态线程，如果对内核线程进行动态管理等；

5) 用户进程管理子系统，用于了解用户态进程创建、执行、切换和结束的动态管理过程，了解在用户态通过系统调用得到内核态的内核服务的过程；

6) 处理器调度子系统，用于理解操作系统的调度过程和调度算法；

7) 同步互斥与进程间通信子系统，了解进程间如何进行信息交换和共享，并了解同步互斥的具体实现以及对系统性能的影响，研究死锁产生的原因，以及如何避免死锁；

8) 文件系统，了解文件系统的实现，与进程管理等的关系，了解缓存对操作系统 IO 访问的性能改进，了解虚拟文件系统（VFS）、buffer cache 和 disk driver 之间的关系。

9) 外设驱动，了解设备驱动程序是如何与外设具体打交道的。这里主要有存储设备，串口设备，以及网络设备等。而且可以进一步了解设备驱动与内核中其他部分的交互与联系。

其中每个开发步骤都可以建立在上一个步骤之上的，就像搭积木，从一个一个小木块，最终搭出来一个小房子。在搭房子的过程中，完成从理解操作系统原理到实践操作系统设计与实现的探索过程。

#### 四、毕业设计方面

对于大学计算机专业方面的本科学生，可以基于 Xv6 或参考 Xv6 进行操作系统类的扩展，形成丰富的毕业设计，比如：

- 进程调度算法
- 内存替换算法和策略

- 文件系统
- 同步互斥机制
- 多核支持
- 缓存机制
- 图形显示支持
- 不同硬件平台的移植
- 实现类似 Linux 的系统调用
- 扩展功能支持不同的应用程序

## 五、总结

Xv6 操作系统的简洁设计和相对全面的内核功能实现，有助于学生学习和掌握操作系统的知识，提高动手实践能力，目前已经被多所高校的操作系统课程采用。就我们目前的实践经验表明，Xv6 适合用于本科计算机专业的操作系统课程的实验，也适合于操作系统相关的本科毕业设计。