https://farid.ps/articles/industrialization_of_profiling/zh.html

从兴趣人物到人口群体：画像的工业化

画像的兴起——从针对特定“兴趣人物”的手动精准审查，到对整个群体进行自动化、持续监控——代表了权力行使、技术角色以及个人自主边界中最深刻的转变之一。曾经需要大量人力努力、机构优先排序和刻意选择的活动，已演变为一种无缝的基础设施，它以日常生活为附带产物，实时生成、聚合并分析数十亿人的行为数据。

这一转变并非单纯由技术所预定。它源于官僚机构的扩张、反复的安全危机、与数据货币化相关的经济激励，以及数据收集、存储和推断的边际成本持续降低。结果不仅仅是“更多监控”，而是一种质上不同的体制：它用无摩擦的规模取代自然的摩擦，用算法自动化取代人为判断，用对多数人的基线观察取代对少数人的例外怀疑。

其核心在于一种根本性的变态：画像从一种**手工艺**——选择性、劳动密集型且解释性的——转变为一种**工业流程**——普遍性、自动化且预测性的。以下内容追溯这一转变，识别那些约束被侵蚀、新能力结晶为持续、全人口推断系统的关键时刻。

I. 基础：作为选择性手动实践的画像

画像在其最基本的形式中，涉及系统地收集和解释信息，以推断特征、预测行为或分配风险类别。其起源可追溯到古代深处。

古代帝国进行人口普查，不仅是为了征税或征兵，也是为了分类。罗马当局和中国帝国行政官员根据职业、忠诚度和地位对人口进行分类，产生了早期的关系地图，能够识别潜在威胁。宗教机构维护出生、婚姻、忏悔和道德行为的记录，构建了揭示影响网络和偏差的原始社会图谱。

然而，这些系统有一个共同的决定性约束：**信息昂贵**。收集、验证、存储和解释数据需要大量人力劳动。因此，画像仍然是**选择性的、偶发的和有界的**。它专注于精英、异见者或战略相关群体——而非整个人口。

在近代早期的欧洲，即使国家扩大了监控 apparatus，这种选择性依然存在。情报工作通过线人、截获信件和物理监视针对异端、政治对手、走私者和外国特工。法国和其他国家的“黑色内阁”（**cabinets noirs**）体现了这种方法：文员团队手动打开信件、复制内容并重新密封以交付。这些操作本质上是受限的。它们专注于高价值目标，因为任何更广泛的行动在后勤上都不可能。

即使在这一阶段，**元数据**的力量也已被清楚理解。关于通信的信息——发送者、接收者、时间和路线——可以在不访问内容的情况下暴露网络和意图。1844年英国邮局间谍丑闻将这一点鲜明地带入公众视野。意大利革命者朱塞佩·马志尼作为伦敦的流亡者，怀疑他的信件被当局应外国势力要求打开。他和支持者在信封内放置罂粟籽和沙粒作为标记；当信件到达时被扰动，马志尼促使激进议员托马斯·邓库姆在议会提出此事。随后爆发的丑闻揭示了内政大臣詹姆斯·格雷厄姆爵士根据令状系统性打开信件的行为，引发愤怒、议会调查，并最终导致邮局秘密部门的废除。这是现代隐私恐慌的早期案例之一，并强调了仅靠关系数据就能瓦解关联网络。

作为回应，“通信保密”（**Briefgeheimnis，secret de la correspondance**）等法律规范应运而生。这些原则将通信数据的使用严格限制在交付等操作目的，禁止用于监控或画像的二次利用。其基本理念简单而深刻：

> 为特定功能生成的数据不应被重新用于构建个人或网络的更广泛画像。

这一原则将在数个世纪中回响——但最终在技术和机构压力下被侵蚀。

II. 官僚世纪：没有自动化的扩展

20世纪极大地扩展了画像，同时保留了许多早期约束。全面战争的需求要求前所未有的信息收集。邮件审查、信号情报和密码破译将监控从精英扩展到更广泛的人口。国家安全局等机构将大规模拦截制度化，而国内机构则编制了关于政治团体、涉嫌激进分子和犯罪网络的广泛档案。

然而，画像仍然**本质上是针对性的**。窃听与特定个人或线路相关。情报档案由人类分析师精心整理。即使数量增加，**人类注意力仍是瓶颈**。

早期计算系统（1950年代–1970年代）开始改变记录保存的规模。政府和企业将福利名册、信用历史和刑事数据库数字化，从而实现更快的检索和交叉引用。但这些系统仍基于**离散记录**，而非连续的行为流。

到1970年代，对集中“数据库”的担忧引发了法律回应。美国1974年《隐私法》和早期欧洲数据保护法引入了目的限制、数据最小化和透明原则。这些框架将通信保密的逻辑扩展到数字时代。

然而，它们建立在一个关键假设之上：数据收集是**有界和偶发的**。它们规范的是记录——而非流动。这一假设很快就会崩溃。

III. 转折点：从记录到数据尾气

决定性的突破发生在1990年代末和2000年代初，随着互联网的兴起——不仅仅作为通信媒介，而是作为持续产生数据的基础设施。

数字系统产生**数据尾气**：作为普通活动副产品自动创建的元数据。每一次连接、查询、点击和移动都会产生可被记录、存储和以微不足道成本分析的痕迹。

这标志着决定性的转变：

> 画像不再是作用于数据的活动，而成为持续产生数据的基础设施。

互联网服务提供商捕获连接日志、DNS查询和路由信息，即使没有内容访问也能揭示行为模式。与邮政元数据——短暂且分散——不同，数字元数据是持久的、集中的且可轻松搜索的。

在这一基础设施之上，谷歌和Meta等平台将画像转变为核心经济模型。搜索引擎捕获意图；社交网络映射关系；移动生态系统追踪移动。嵌入式追踪器将可见性扩展到网络的广大部分。Meta的追踪像素存在于全球约三分之一的热门网站上，监控远超出其自身平台的活动，常常捕获来自健康、金融或政治背景的敏感信号。

在这种环境中，一个关键认识浮现：

> 内容在很大程度上变得多余。在许多情况下，关系模式不仅是意义的代理——它们在分析上比内容本身更有用。

元数据不仅仅表明通信发生；它使**内容的概率重建**成为可能。谁与谁通信、何时、多频繁以及在何种更广泛的背景下，可以强烈限制正在通信的内容。公开可用的信息——共享的从属关系、专业角色、政治立场、社会联系——进一步缩小了可信解释的空间。

随着时间推移，这些约束变得具有预测性。元数据不仅是描述性的；它是生成性的。它不仅仅伴随内容——它常常能**近似或推断**内容，尤其是在大规模聚合时。

搜索查询揭示意图。通信频率揭示关系强度。共同位置揭示关联。在足够规模下，这些信号汇聚成高度准确的行为模型，通常使直接内容访问变得不必要。

企业系统为货币化优化行为；国家系统为控制约束行为——但两者都依赖相同的底层机制：**通过大规模行为推断进行预测**。

IV. 没有逃脱的身份：持久锚点

工业画像的一个定义性特征是**持久身份**的出现。

早期系统依赖可变的标识符——姓名、文件、地址——这些可以被更改或掩盖。现代系统通过重叠信号重建身份：

- 设备指纹
- 行为模式
- 社交图谱
- 生物识别标记（面部、步态、声音）

公开共享的图像作为持久锚点。即使个人更改账户或采用化名，面部识别系统——尤其是在国家或情报背景下——也能跨数据集重新连接身份。照片或共享事件中的共同出现进一步强化推断的关系。

其含义深刻：

> 身份不再是某人声明的东西，而是持续推断的东西。

这消除了曾经约束监控的大量摩擦。识别不再依赖单一信号；它从众多信号的冗余中浮现。

V. 融合：从数据点到本体论

这一演变的顶点是**数据融合**：将分散的数据集整合到统一的分析系统中。

Palantir Technologies等平台将政府记录、金融交易、社交媒体活动、位置数据和通信元数据聚合到个人和网络的连贯模型中。这些系统构建动态本体论，允许分析师查询关系、检测模式并生

成预测。

一个具体例子说明了这一转变。在移民执法中，Palantir的“增强线索识别和执法针对”（ELITE）工具用潜在目标填充地图，借助签证记录、就业数据、电话元数据、社会联系，甚至医疗补助或HHS地址信息来分配“地址置信度分数”并生成档案。官员可以识别“目标丰富”的社区进行行动，标记个人不仅基于直接证据，而是因为他们的**行为和关系签名**类似于先前识别的案例。类似融合出现在ImmigrationOS等工具中，该工具整合旅行历史、生物识别和社会数据以进行优先排序。

怀疑不再是被发现的——而是被**生成的**。

画像不仅仅记录现实；它通过浮现概率关联主动构建现实，这些关联成为可操作的。

VI. 从解释到预先阻止

传统画像主要是回顾性的。它寻求解释过去的行为——谁犯了罪，谁组织了阴谋，谁构成了威胁。

工业画像是预测性和预先阻止性的。它识别：

- 谁可能犯罪
- 犯罪可能在哪里发生
- 谁可能违约、激进化或偏离

这一逻辑常被比作《少数派报告》中描绘的愿景，在那里个人在犯罪前被逮捕。虽然当代系统缺乏确定性的预见，但结构相似性很明显：预测性警务工具分析历史数据、911呼叫、车牌阅读器和社会信号，以生成“热点列表”或风险分数。

现代系统基于概率运作。个人被标记不是因为他们会行动，而是因为他们**统计上类似于那些已经行动的人**。

这一转变微妙但深刻：

> 个人不再主要根据其行为被判断，而是根据其在概率景观中的位置。

怀疑成为结构性的——持续生成而非由离散事件触发。

VII. 推断时代的法律

《通用数据保护条例》等法律框架试图通过同意、透明和最小化来施加限制。然而，它们面临结构性约束。

大多数法律系统规范**作为对象的数据**。现代画像的力量源于**关系和推断**，这些远更难定义、观察或约束。

额外的挑战包括：

- 跨司法管辖区的持续数据流
- 国家安全和“合法利益”的广泛例外
- 对监督有抵抗力的不透明算法系统

结果是持续的不匹配：

> 为记录时代设计法律框架难以治理持续、预测性推断的时代。

VIII. 权力的不对称

工业画像产生结构性不平衡。

个人通过参与现代生活持续生成数据。避免是可能的，但代价高昂且不完整。与此同时：

- 企业维持受保密保护的不透明系统
- 国家通过法律权威或伙伴关系访问和整合数据
- 技术复杂性掩盖问责

结果是不对称的清晰：

> 多数人变得可读；强者相对保持不透明。

IX. 内化：画像与行为的自我调节

超越其制度和技术维度，画像的工业化产生深刻的心理转变。监控不再仅作为外部力量运作；它被内化。

米歇尔·福柯在其对全景监狱的分析中预见了这一动态：杰里米·边沁的理论监狱设计，其中囚犯对他们看不到的中央观察者可见，从而内化纪律并在持续监视的不确定性下自我调节。全景监狱的力量不在于永久观察，而在于对其的**预期**。

工业画像极大地扩展了这一逻辑。个人在环境中运作，其行为可能以不透明方式被记录、分析和解释——由优化参与的平台或评估风险的国家。结果是向**自我调节**的转变。

这表现为：

- 在帖子、搜索或关联中的自我审查
- 避免某些群体、话题或地点
- 与感知规范对齐以最小化风险分数
- 在数字和物理语境中修改行为

关键是，这些适应不需要明确胁迫。它们源于预期。

> 控制不仅通过系统所做的行使，还通过个人避免做的行使。

影响延伸到个人之外。随着人们自我审查和自我分类，生成的数据强化模式，塑造未来的预测。系统不仅观察现实——它微妙地重塑现实，创造强化顺从的反馈循环。

X. 选择性监控的终结

画像经历了根本转变：

- 从**针对性**到**普遍性**
- 从**手动**到**自动化**
- 从**回顾性**到**预测性**
- 从**碎片化**到**整合化**

早期系统受摩擦约束——成本、时间、人类注意力。工业系统移除这些约束。监控成为环境性的。包含成为默认。

“数据应仅服务其直接目的”的原则已让位于一种范式，其中**所有数据都可能被利用**。

XI. 结论：参与的代价

从邮政保密到数字数据融合的漫长弧线揭示了一个一致模式：每次技术扩张都增加画像的范围，而法律和社会回应滞后于其后。区分当下的是结构性的。画像不再是针对特定个人的活动——它是个人存在于其中的基础设施。

“兴趣人物”的类别溶解。每个人都成为持续评估的对象。

这一转变不仅由国家权力维持，还由经济激励维持。看似免费的平台通过行为数据提取运作。“如果你不付费，你就是产品”这句话捕捉了一种直觉——但低估了现实。

被生产的不是个人，而是个人的**预测模型**——可移植、可操作，且往往对它所代表的人不可及。

一个中心挑战在于感知与现实之间的差距。

首先，人们低估了已知内容的**影响**。画像通过关联运作。关系——过去的、弱的或间接的——可以塑造结果。与后来变得不受欢迎的人的联系可能影响机会。一个人不仅作为个体被判断，还作为关系被判断。

其次，人们低估了可知内容的**范围**。系统从模式而非明确披露推断敏感属性——政治的、宗教的、性的、经济的。这些推断无论准确性如何都变得可操作。

> 个人不仅根据他们揭示的内容被评估，还根据可推断的内容——以及他们与谁相关联。

数字生活的参与因此涉及隐含交换：便利换取可读性。这一交换既不透明也不可协商。

挑战不是停止数据化，而是约束它——恢复摩擦、强制限制并确保问责。

中心问题很清楚：

> 干预是否会在永久画像的基础设施变得过于根深蒂固而难以有意义挑战之前发生？

若无此类干预，参与的代价将不仅仅是数据——而是观察、推断并最终被定义之间的边界逐渐侵蚀。