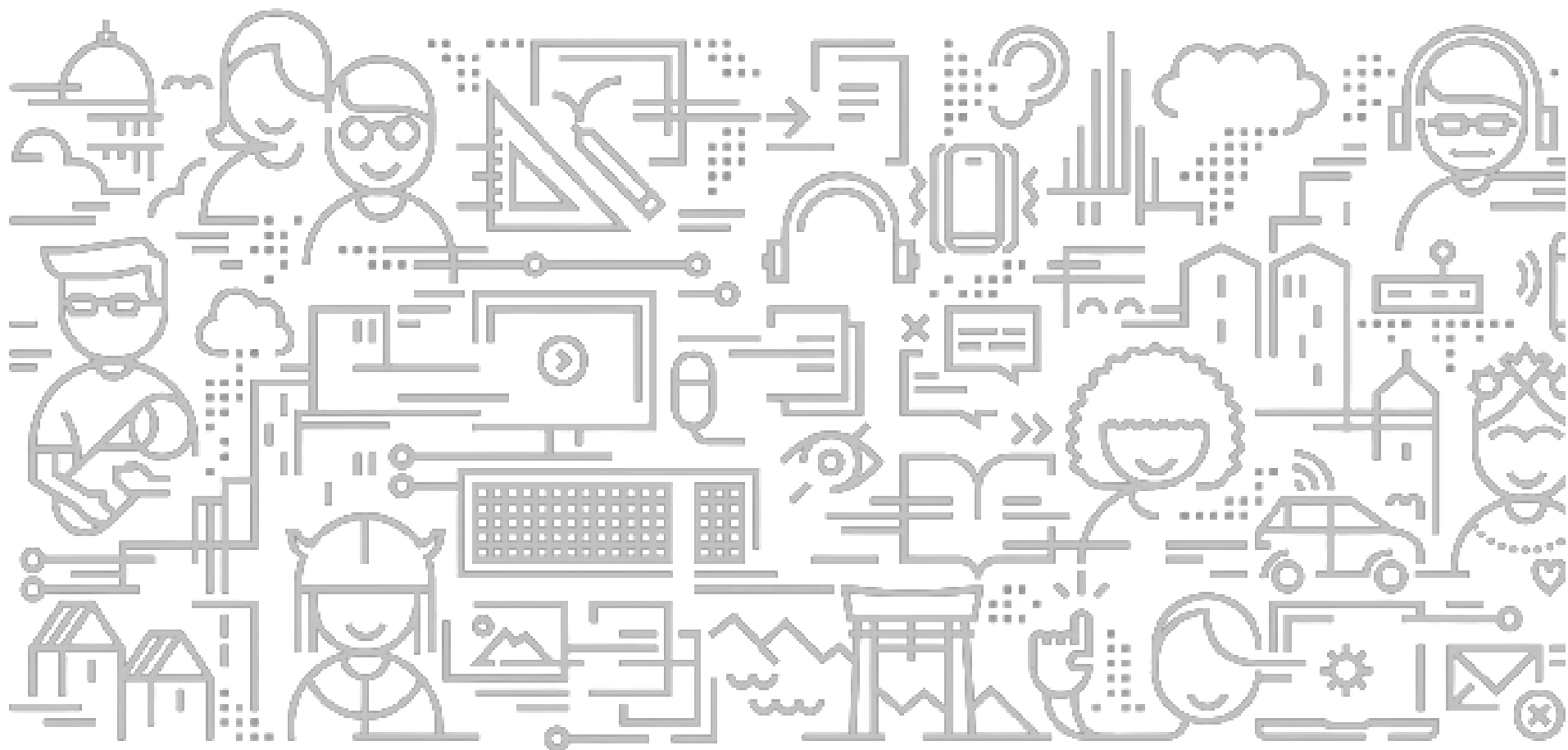


# 包容性设计原则手册



上海有人公益基金会 微软公司 小米集团

编辑：傅高山 张开羽

# 前言

被产品/服务设计排除在外是什么感觉？残障人士已经有过太多与之相关的经历。公共交通是现代城市交通系统的重要组成部分，但忽略无障碍设计的公交车却把轮椅使用者挡在门外；智能手机的出现使人与人之间的联系比以往任何时候都更密切，但它最初的设计中不包含辅助视障群体操作的屏幕朗读系统；“数字互联能够改变教育”的观念引起了全球的关注，但残障儿童尤其是心智障碍儿童依然缺少他们能够使用并乐在其中的数字产品。

残障人士遭遇排斥，社会与企业也因此蒙受损失。残疾人权利公约秘书处 (SCRPD)在其发布的关于残障人士的事实清单(<https://www.un.org/development/desa/disabilities/resources/factsheet-on-persons-with-disabilities.html>)中援引英国科技公司 Nomensa 于 2006 年实施的一项研究指出，英国伦敦证券交易所富时 100 指数所纳入的企业，有 75%未达到网页无障碍的基本标准，从而错失超过一亿四千七百万美元的收入。

上个世纪 50 年代以来，人们对残障的认知模式不断更新。残障不再被认为是个人的悲剧与缺陷，而被视为社会建构的结果。过往社会规范的制定、社会环境的建设完全以“健全人”为原型，忽视了残障人的差异性需要。因此，解决残障问题的关键并非是在医学意义上治愈残障人士肢体、视力、听力等方面的功能“缺失”，而是在尊重的基础上，将残障人的差异性需要融入社会规范，重建对该群体成员开放、融合、友好的社会环境。在这一背景下，无障碍(accessibility)的概念应运而生，并成为国际社会的共识，各国纷纷以立法的方式为无障碍提供制度性保障。

随着无障碍实践的发展，人们逐渐意识到，无障碍产品与服务的受益者不只是残障人，还有身处各类障碍情境中的非残障人。20 世纪 70 年代，“为残障人设计”开始向“为所有人设计”转向。“残障”在设计领域被重新定义为个人需求与产品、服务、环境的设计不匹配。任何人都可能在某一时期体验到残障。无障碍并非某一个群体的专属需求，而是全体社会成员的共同需要；无障碍也并非社会为了照顾个别成员所做的额外付出，而是引领社会建设与技术发展的重要航标与创新源泉。

在很长一段时间内，无障碍的概念仅限于定义物理空间。然而如今，数字技术正在强有力地介入我们与世界的互动过程。应用程序、网页和更多面向公众的技术产品的设计者们有责任尽可能地减少设计失败导致的人-技术-世界之间的障碍。数字时代带来了许多思想与范式的转变——但只有当每个人都能参与，这些转变才是真正的进步。为促进包括残障人士在内的全体社会成员充分、切实地融入社会生活的方方面面，也为给开始关注无障碍的人们提供理念与实践指导，支持其在数字包容领域的探索，上海有人公益基金会在微软中国和小米集团的支持下，编写了这本手册。

本手册以微软包容性设计原则手册为蓝本(<https://www.microsoft.com/design/inclusive/>)，并结合小米集团自 2016 年以来在智能手机、AIoT 等方面的无障碍与包容性设计实践，以及上海有人公益基金会发起方一加一残障公益集团自 2006 年以来在中国的无障碍倡导经验。上篇综合阐述无障碍与包容性设计的理念与原则，下篇就人工智能的包容性以及如何避免偏见进行了专题分析。我们期望本手册的出现，能够引发更多关于无障碍与包容性设计的讨论。在思想观念领域构建无障碍，再将之不断地落实到现实中。期望每一个产品的设计，都能助力每一个生命，使之拥有更多可能。

手册上篇源自微软公司《包容性设计指引》(Inclusive 101)，它凝聚了微软及其他团体的见解与洞见，其中有数位关键人物需要特别提及：朱塔·特雷维拉努斯(Jutta Treviranus)，安大略艺术设计大学包容性设计负责人，包容性设计研究中心和包容性设计学院的领头人、创始人；格雷厄姆·普林

(Graham Pullin), 邓迪大学数字交互设计与产品设计教授, 《设计遇见残障》(Design Meets Disability)一书的作者; 艾伦·萨耶格(Allen Sayegh), 哈佛大学设计研究生院建筑技术实践副教授, INVIVIA 设计公司负责人。此外, 上篇结尾部分提供的工具包由阿尔伯特·岑(Albert Shum), 凯特·福尔摩斯(Kat Holmes), 克里斯·伍勒里(Kris Woolery), 玛格丽特·普莱斯(Margaret Price), 道格·金(Doug Kim), 埃琳娜·德沃基纳(Elena Dvorkina), 德里克·迪特里希·穆勒(Derek Dietrich-Muller), 内森·凯尔(Nathan Kile), 莎拉·莫里斯(Sarah Morris), 乔伊斯·周(Joyce Chou), 索戈尔·马利克扎德(Sogol Malekzadeh)以及致力于推进包容性设计的微软团队制作。

手册下篇源自微软公司包容设计团队乔伊斯·周(Joyce Chou), 罗杰·伊巴尔(Roger Ibars), 奥斯卡·穆里略(Oscar Murillo)撰写的《追求包容性人工智能》(In Pursuit of Inclusive AI)。其他贡献者有丹妮尔·麦克卢恩(Danielle McClune), 道格·金(Doug Kim), 埃琳娜·德沃基娜(Elena Dvorkina)、伊兹·贝尔特兰(Izzy Beltran)和蒂芙尼·陈(Tiffany Chen)。

# 目录

## 上篇

关于包容性设计的事实.....	6
包容性设计的原则 .....	10
识别排斥.....	10
从多样性中学习.....	14
适用于一，到适用于多.....	17
从观念到行动.....	24
工具包.....	27

## 下篇

识别人工智能的偏见.....	29
创造包容性人工智能.....	37
附录：联合国可持续发展目标实施背景下 《关于科技产品遵守信息无障碍相关原则的 北京倡议》 .....	44

# 关于包容性设计的事实

事实上，作为设计师，我们通常基于我们已有的知识来生产和评估想法。我们努力创造能够满足需求，切合人体结构，提高生活品质的体验。

但问题在于：如果我们以自己的能力为基线，那么我们创造的产品必然只便于一部分人使用，而对其他人并不友好。比如老人、女性、残障人等。

全球共有 74 亿人。我们的目标是制造在身体、认知和情感上适用于每一个人的产品。这个目标的实现从将人类多样性视作更优设计的资源开始。

## 我们为谁设计？

如果我们以个人的能力和偏见作为起点，最终，我们仅仅为拥有某一性别、年龄、语言能力、技术素养以及身体能力的人群设计产品。我们所假定的客户群体，能够以某种特定的方式获取金钱、时间和社交网络。

比如，当下以视觉为中心的设计，常常会将视障人士排斥在外。

## 谁被排斥了？

当谈论人时，所谓“正常”的概念是不存在的。我们设计的交互很大程度上取决于我们自身的所见所闻、所感所知。

所有感官和能力一直都处于最大化状态的假设，使我们损失了很大一部分关于人的现实处境的视野。

## 包容性设计为什么重要？

基于无障碍与包容性原则的设计不仅扩大了我们的产品和体验的受众范围，还反映出了人类的真实面貌。所有人都在不断地成长、改变，适应周围的世界。我们希望我们的设计能够反映出这一多样性。

我们所做的每一个决定都有可能抬高或者降低社会参与的门槛。而通过无障碍产品、服务、环境以及体验来降低这一门槛，让全世界更多的人享受科技带来的便利，是我们共同的责任。



## 包容性设计的定义

包容性设计：一种能够充分利用并实现人类多样性的设计方法。

最重要的是，包容性设计意味着接纳拥有不同观点的人，并向他们学习。

包容性设计并不是设计一个适用于所有人的东西。你在设计多样化的方式，以使每一个人都能有归属感地参与其中。

很多人无法参与到社会的某些方面，无论是物理意义上的，还是数字生活。理解人们为什么以及如何被排斥的，为我们实现包容性设计提供了可操作的步骤。

## 无障碍的定义

《在标准中界定无障碍的指南》（第二版）将无障碍（Accessibility）定义为“产品、服务、环境和设施能在多大程度上被最大范围的不同特征和能力的人群使用，以在特定使用环境中实现特定目标”。

从宏观上说，无障碍包括人文环境无障碍、物质环境无障碍、信息和沟通无障碍三大类。无障碍既强调普遍服务，又重视特别支持。

无障碍是一种属性，而包容性设计是一种方法。虽然实践包容性设计能够提高产品的无障碍，但这不是一个力图满足所有无障碍标准的过程。在理想情况下，无障碍和包容性设计共同创造出不仅符合标准，而且真正可用，并向所有人开放的体验。

### 识别排斥

#### 当我们用自身的偏见解决问题，排斥就会发生

几乎从世界卫生组织 1980 年首次发布其对"disability"的正式定义的那一刻起，我们对“残障”与“限制”的认识就在不断地发展变化。2001 年发布的一项新定义关注到了人与社会的相互作用。如今，当我们谈论残障和与之相关的限制，我们也在谈特定情境下的障碍、活动限制，以及社会参与受限。我们将个人与环境、情景和社会整体之间的不匹配都囊括其中。

1980

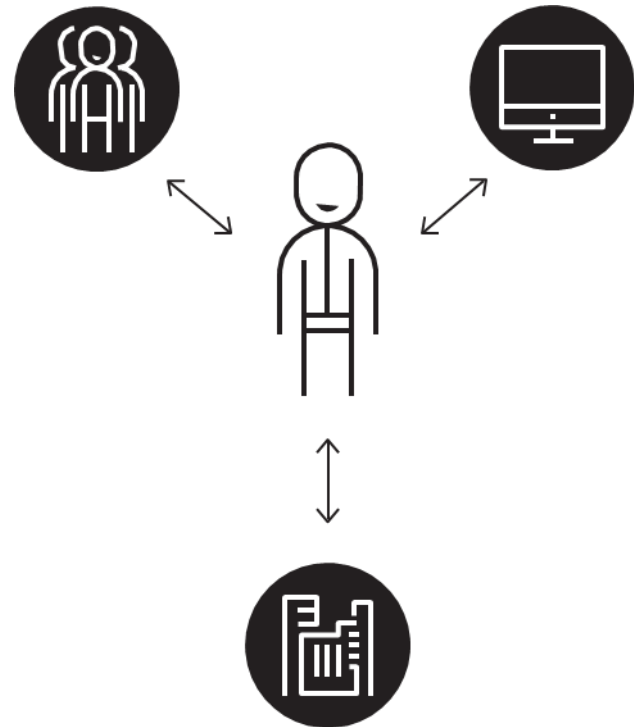


## 残障是个人特征

“在健康体验的语境中，残疾指的是，以人类看来正常的方式，或者在正常范围内从事活动，（由于损伤）所受到的任何限制或能力缺失。”

-世界卫生组织

如今



## 残障取决于环境

“残障不只是一个健康问题。它是一个复杂的现象，反映出个人的身体特征与其所生活的社会的特征之间的互动。”

-世界卫生组织

残障发生于个人与社会的交互点。生理、认知和社会排斥是不适配的交互的结果。比如，当视障者遭遇图形验证码，听障者遭遇语音沟通，动作不灵便者遭遇一闪而过的悬浮窗口，他们就处于残障的状态。反之，如果在设计时考虑多样性，提供替代方案、通用方案，如短信验证、电话验证、文字沟通等，他们就不再“残障”了。

作为设计师，我们有责任了解我们的设计如何影响人与社会的交互，如何导致错配。

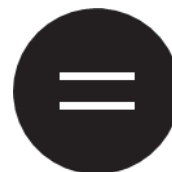
理解排斥有助于我们生产出新的想法与包容性设计。这显明我们有机会为最多数人提供实用而优雅的解决方案。

**残障**



**个人健康状况**

**残障**



**不适配的交互**

## **有时候，排斥是暂时的**

短期的损伤或者环境也会影响人们与周围世界的互动方式，即使这种影响是短暂的。请设想以下情况：注视明亮的灯光，打着石膏，或者在异国他乡预订晚餐。

## **有时候，排斥是情境性的**

当人们进入不同的环境中，他们的能力也可能大幅改变。比如，在嘈杂的人群中，人们听不清楚；在汽车里，人们有视觉盲区；新晋父母每天大部分时间都在单手做任务；长时间高强度工作会导致官能超负荷。什么是可能的、安全的、合适的？答案在不断变化。

因此，我们需要理解，每个人都有可能是“残障者”。无障碍与包容性设计关乎所有人。

# 从多样性中学习

## 人类是适应多样性的真正专家

包容性设计从一开始就以人为中心。你需要全新的、多样化的观点，使包容性设计成为可能。

人类具有适应不同场景的惊人能力。比如，不用重新学走路，儿童也能够让自己的步法快速适应从平地到上坡的转变。不同的人与不同的环境相适应的过程与方法，本身也成为多样性的体现。而理解这些适应力是获取真知灼见的关键。

**更多的技术变动**



**更多的障碍时刻**

人机交互很大程度上取决于我们可以看见、听见、说出、触摸、学习、记忆什么。如今日新月异的技术使得情境性障碍与许多人高度相关。

人们往往不得不以改变他们的习惯为代价，接受新的情景。

技术的多变向我们提出了这样的问题：是我们被迫适应技术，还是技术适应我们？

## 理解适应的过程

同理心在许多不同的设计形式中都很重要。当建立对社会排斥和残障的同理心时，仅仅依靠眼罩和耳塞来模拟不同能力人群的体验是具有误导性的。学习人们如何适应周围的世界，意味着花时间从他们的角度理解他们的经历。

我们可以尝试想象一个拥有一系列特定能力的人将如何使用某种产品或者服务，但是我们无法知晓他使用时的情绪，什么让他欢乐，什么让他沮丧。所以，重要的不是自己想象，我们可以邀请对方参与进来，尽情地分享他们的奇思妙想，和我们一起来设计。在这个过程中，设计师也是受益人。

如果做得好，我们不仅可以了解人们遇到的障碍，还能了解所有人共享的动机。



## 适用于一，到适用于多

### 专注于对所有人普遍重要的事情

人们有一些普遍的体验世界的方法。所有人都有动机，并建立关系；都有能力，而这些能力是有限的；与设计交互时，每个人都会经历到一些排斥。

另一方面，一个对盲人有效的解决方案也能使汽车驾驶员受益。比如，定速巡航系统的发明者就是一位视障者，他想发明这个装置的原因是坐在车里和司机聊天时忽快忽慢的车速让他很难受。而这一装置一开始能够普及并非是因为它解放了司机的右脚，而是 1973 年爆发的中东石油危机，让它凭借省油的优势脱颖而出。

包容性设计跨越了一系列相关能力，将相似情况下不同的人联系起来。



**听力受损**



**阅读机场字幕**



**教孩子阅读**

## 限制之美

为长期处于残障状态中的人设计似乎面临投入多但受益人少的巨大限制，但多项实践已经证明，由此产生的设计结果实际上可以带来更为广泛的积极影响。例如，为听障群体设计的隐藏式字幕，也为其他人带来了好处：你可以在喧嚣的机场阅读字幕，或者更好地教孩子阅读。

类似地，高对比度屏幕设置最初是为视力障碍人群设计的。但如今，在强光下使用屏幕的人也能从中受益。远程控制、自动开门装置、有声读物、电子邮件等也是如此。纳入障碍因素的设计，就是优秀的设计。



**独臂**



**手臂受伤**



**新晋父母**

## 使不同的人群受益

为长期处于残障状态中的人设计，也能造福特定情境下有障碍的人群。例如，一个为独臂人士设计的装置，对手腕受伤的人，或抱着婴儿的新晋父母可能同样有用。我们将这称为“用户画像频谱”。



**永久障碍 26000**



**暂时障碍 1300 万**



**情境性障碍 800 万**

**共计 2100 万以上**

## 使更广泛的人群受益

认识到长期性障碍和情境性障碍是一个连续统一体，有助于我们重新思考如何以新的方式将设计推广到更多人。在美国，每年有 26,000 人遭受失去上肢的变故，若将有暂时和情境性障碍的人包括在内，这个数字将超过 2,000 万。

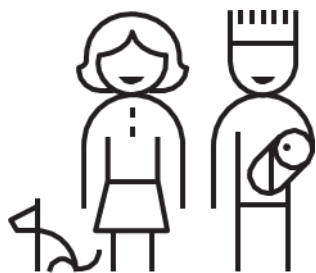
*资料来源：美国人口普查局，肢体生命基金会，截肢者联盟，MedicineHealth.com，CDC.gov，加州大学旧金山分校残障数据中心*

	<b>永久障碍</b>	<b>暂时障碍</b>	<b>情境障碍</b>
<b>触</b>			
	<b>独臂</b>	<b>手臂受伤</b>	<b>新晋父母</b>
<b>视</b>			
	<b>盲人</b>	<b>白内障</b>	<b>分神司机</b>
<b>听</b>			
	<b>聋人</b>	<b>耳部感染</b>	<b>酒吧侍者</b>
<b>说</b>			
	<b>永久失声</b>	<b>咽炎</b>	<b>重口音</b>

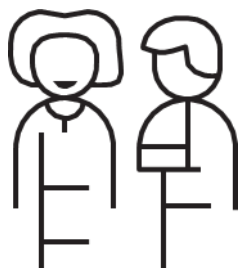
## 用户画像频谱

我们使用用户画像频谱来理解长期障碍、暂时障碍、情境性障碍中的错配与动机。

如左图所示，用户画像频谱是一个可用于培养同理心，并展示解决方案如何推广至更广泛受众的快速工具。



家人



朋友



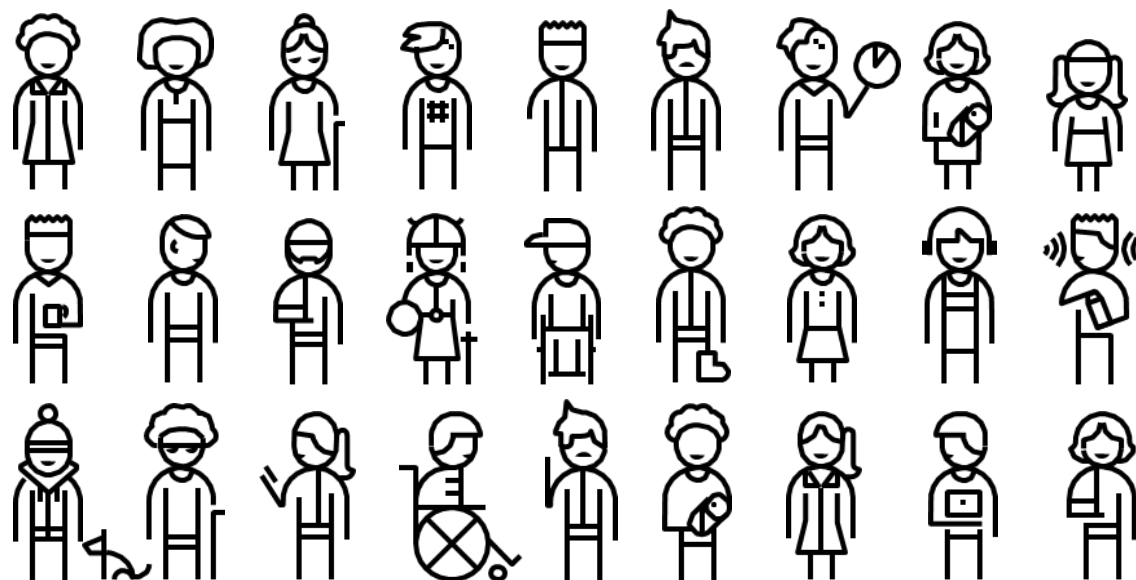
陌生人

## 人物关系网

没有人是一座孤岛，用户画像频谱也不是孤立的。

我们的人物关系网囊括了朋友、同事、家庭成员，甚至是陌生人。

我们致力于建立能够团结大众并包容个性的体验。这些体验根植于我们与周围世界的人或物的交互。以不同的方式看待残障并理解排斥，有助于我们将适用于一个人的解决方案拓展为适用于成千上万人的解决方案。



## 从观念到行动

我们将包容性设计视为一系列高举人类多样性的观点和实践。请花一点时间思考，你如何从原本为残障人士设计的解决方案中受益。请留意你自身的长期的、暂时的和情境性障碍对你与社会交互的阻碍。观察不同情况中的人们如何被排斥在那些你可以无障碍地完成的事情之外。

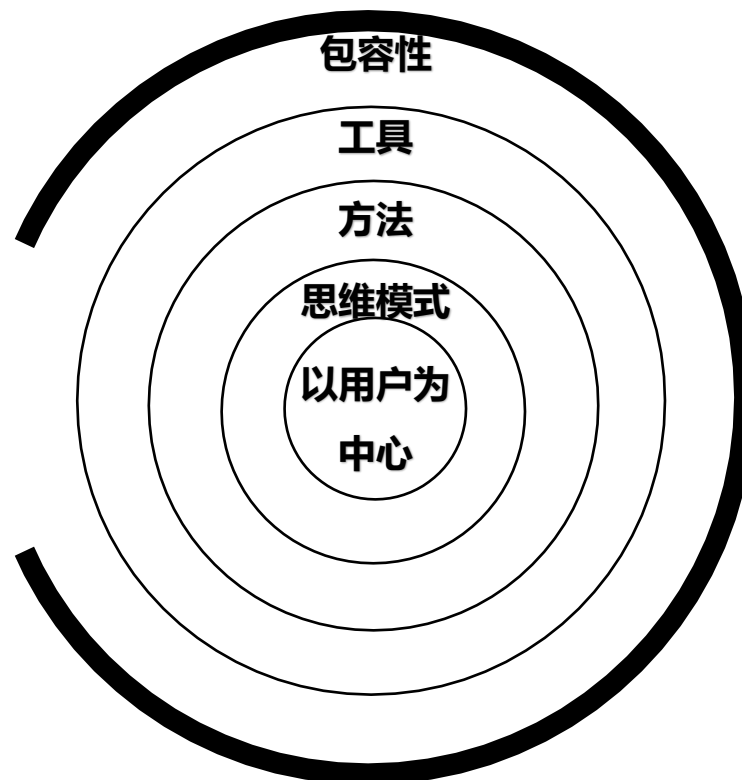
改变观念是走向融合的第一步。

现在是时候将观念付诸行动，打造新的设计方案了。让我们开始吧。



传统的以用户为中心的设计运用许多技术来界定人们的需求：从用户画像到使用场景，再到可用性测试。但是我们还需要能够将多样性重新引入设计流程的工具。

我们需要方法，来检查、平衡和衡量设计的包容性。



## 包容性设计的益处

基于无障碍与包容原则设计的技术将带来以下益处：

1. 更易使用
2. 更少分歧
3. 更富情感

包容性设计不仅影响了人们使用的产品，还会转变我们的观念、方法和行为。我们设计的内容是我们设计方法的副产品。衡量其收益包括衡量我们的文化和我们自身的转变。

## 工具包

这套指引是微软包容性设计工具包的一部分。该工具包可在现有设计流程中运用。它基于以下三个原则：

1. 识别排斥
2. 从多样性中学习
3. 适用于一，到适用于多

我们可以使用这个工具包来评估我们现有的流程，并发展新的实践。随着实践经验的积累，工具包将不断完善。

## 实际应用

大多数设计流程都是迭代和启发式的。包容性设计工具包旨在完善，而不是取代现有的许多设计流程类型。我们可以从多种途径获取以人为中心的设计方法。就像厨师的菜谱一样，你自己的设计流程应该是你设计的基本方向；你可以添加工具包内的元素，像是添加食材，以提高你的设计流程的包容性。如何以及何时添加取决于你。

## 更多资源

请跳转至 [microsoft.com/design/inclusive](https://microsoft.com/design/inclusive)

1. 活动：下载 PDF 文件“包容性设计活动”（Inclusive Activities），了解如何在设计的不同阶段更加关注包容性。
2. 视频：你可以在这里观看多部短片，以及一部关于包容性设计的 20 分钟的纪录片。

作为设计师，我们有责任了解我们设计的交互的力量。拥抱人性的设计，驱使我们创造更美好的世界，也带来了更具包容性的技术。

欲了解更多信息，请联系：[inclusivedesign@microsoft.com](mailto:inclusivedesign@microsoft.com)

# 识别人工智能的偏见

为在设计阶段就将残障人士纳入考虑，微软开发了包容性设计工具和流程（见上篇）。随着实践的发展，微软将包容性设计理念拓展到了其他受排斥影响的领域，包括认知问题、学习方式偏好和社会偏见。

如果人工智能在设计之初忽视了包容性，偏见同样会出现。已有的经验提醒我们，要创造具有包容性的人工智能，最重要的一步是，意识到设计者的局限，明白哪里有偏见，以及偏见如何影响系统运作。

本手册将人工智能偏见划分为不同的类型，帮助产品开发者尽早发现已有的问题，预测未来可能出现的问题，作出更好的决策。开发团队将清楚地看到他们的系统可能存在怎样的偏差，识别偏见、优化体验，兑现“人工智能适用于所有人”的承诺。

## 识别偏见的五种思路

通过与学术界以及行业思想领袖们合作，微软总结了识别人工智能偏见的五种思路，并且以童年场景作比，说明每种偏见类型对应的行为。为什么这样作比呢？我们都很容易理解这些场景，因为我们或多或少有过类似的经历；而且，童年的比喻非常贴切：人工智能正处在发展初期，未来它如何成长，反映出现在我们如何培养它。

每种偏见类别之下，都包含了相应的定义、依托童年比喻作出的阐释、产品案例，以及一项针对人工智能设计团队与设计工作的压力测试。我们对以下内容进行了一定的本土化处理，以使得案例更贴合中国文化与实践。

五种偏见类型分别是：数据集偏见、联想偏见、自动化偏见、互动偏见、确认偏见。

## 1. 数据集偏见

用于训练机器学习模型的数据不能代表客户群的多样性。大规模数据集是人工智能的基础。但同时，数据集往往被简化为大众版本，不考虑各种各样的用户，因此缺乏代表性。比如，人脸识别数据集可能缺少颜面烧伤者的、老年人的图像。

儿童仅仅依赖他们非常有限的见识定义世界。最终，孩子们会了解到，世界如此巨大，远远超越了他们视野所及的一小部分。这就是数据集偏见的根源：智能的发展基于太少或者太过同质化的信息。

### 产品案例

计算机视觉技术在服务于白人时表现更佳，这是因为最初的训练数据里排除了其他种族与肤色的人群。

### 压力测试

你正在使用的数据集样本，是否代表了客户群中的每一个个体？如果没有，你是否对在样本之外的人群进行了测试？你的人工智能团队成员是否包容、多元、对偏见敏感？

## 2. 联想偏见

用于训练模型的数据加深甚至成倍地放大了文化偏见。对不同性别、残障、民族、种族、肤色、职业等群体的偏见，植根于我们的意识，见诸于日常生活的各种表达。在训练人工智能算法时，人类的偏见会影响机器学习。在未来的交互中延续那些偏见，可能会导致不公平的客户体验。

想象一下那些喜欢玩职业扮演游戏的儿童。男孩们想当医生，并且假定护士该由女孩们扮演。女孩们必须据理力争来推翻这种假定。

“喂，女孩也可以当医生！”

### 产品案例

带有性别偏见的翻译工具（例如，飞行员被假定为男性，空乘则是女性）。

### 压力测试

你的人工智能是否进行了暗含性别或民族刻板印象的联想？为了打破不受欢迎的、不公平的联想，你能做些什么？你的数据集是否已经被分类标记了？



### 3. 自动化偏见

自动化决策藐视社会与文化因素。如今，自动化决策不仅在私人部门和商业活动中被广泛使用，还深度参与到公共部门的政策制定之中。预测程序可能自动违背人类的多样性目标。算法不需要对人类负责，但它们的决定却对人有影响力。人工智能设计师和从业者需要考虑被他们建立的系统所影响的人群的目标是什么。

想象一个女孩正在化妆。这个女孩喜欢运动，喜欢自然的外表，讨厌一切做作的东西。但化妆师对美有不同的看法。他给女孩化了一个浓妆，做了一个复杂的发型。最后，化妆师很高兴，女孩却被自己的装扮吓坏了。

#### 产品案例

美颜滤镜对人脸图像的调节，强化了欧洲人的审美观，例如，他们认为更亮的肤色是更美的。

#### 压力测试

现实中的、多样化的客户都会同意你的算法结论吗？你的人工智能系统是否拒绝人类做出的决定，而倾向于自动化的决策？你如何确保反馈循环中包含人的想法？

## 4. 互动偏见

人为篡改人工智能，造成有偏见的结果。如今的聊天机器人会开玩笑，很多时候还会让人错把它们当成人类同伴。但许多将人工智能人性化的尝试，都在无意中让计算机程序沾染到人类的偏见。比如，有人恶意地宣扬“同性恋是一种病态，需要电击治疗”。在机器人动态学习的过程中，如果对这些偏见不设防，互动偏见就会出现。

“传话游戏”在儿童群体中非常流行：第一个小朋友对下一个小朋友说一句话悄悄话，下一个小朋友再对下一个小朋友重复这句话——如此依次传递，由最后一个小朋友说出他听到的话。游戏有趣的地方是，我们可以观察，经过多次传递，信息会如何变化。然而，如果其中一个孩子故意将他听到的话篡改成另一个荒谬的版本，结果可能更好笑，但是游戏的内核却被破坏了。

### 产品案例

人类故意将种族或者性别歧视的语言输入聊天机器人，训练它们说冒犯性的话。

### 压力测试

你是否纳入了审查机制，以识别那些针对系统的恶意行为？你的人工智能系统从人类身上学到了什么？你设计的是实时互动和学习的人工智能吗？如果是，这意味着人工智能会给用户什么样的反馈？

## 5. 确认偏见

过度简化的个性化设计对群体或个人做出带偏见的假设。确认偏见以先入为主的方式解释信息。人工智能算法总是提供与他人已经作出的选择相匹配的内容，这就排除了那些不流行的选择。类似地，学者如果只从想法一致的人那里获取信息，就永远看不到反对观点，也看不到替代选项和多样化的思路。

假设一个孩子某一年收到了一个玩具恐龙的礼物。其他家庭成员看到了恐龙，于是送给他更多的恐龙。几年后，亲友们就会觉得这个孩子是一个恐龙迷，继续不断地送他恐龙，直到他拥有了数量巨大的同类玩具。

### 产品案例

购物网站向客户推荐他们曾经搜索或者已经购买的商品。

### 压力测试

你的算法是否建立在流行的选项之上，并且强化了这些选择？随着时间推移，你的人工智能系统能依据用户的改变做出动态演算吗？你的人工智能系统能让客户以更多样和包容的视角看世界吗？

## 从这本手册开始

作为人工智能的设计者和创造者，我们有责任思考人工智能如何进化，如何影响人类。让人工智能平等地服务于每一个人，是一条漫长的道路，希望这本入门手册能成为一个起点。

美国曾发生面部识别软件对非裔美国女孩产生误读的事件。在了解了上述内容之后，我们可以把这个案例总结为数据集偏见：该软件使用的训练数据太狭隘了。如果一开始就意识到这个偏见，我们可以纳入其他人为因素对系统进行测试，创造更具包容性的体验。除此之外，我们的面部识别软件会不会被有意为之的错误数据损害？还有哪些偏见会影响我们的体验？

大多数人工智能领域的工作者都听说或经历过这些情形。比如，联合国教科文组织在 2019 年出版的报告《如果我，我会脸红》中指出，多数人工智能语音助手都将声音默认设置为女声，并强化了女性顺从的形象。无意的偏见会导致尴尬的、冒犯的结果。这是我们都想要识别和避免的。我们的目标是帮助你认识这些不利结果背后的潜在偏见。留心这些偏见，从这些偏见类别开始测试你的人工智能体验，这样你就可以专注于向所有客户展现人工智能的潜力。

## 创造包容性人工智能

如今，机器一般从信号和样本中学习。复杂的程序表现得与人类很相像。比如，我们能在聚会上认出熟人；即使几年没骑车，重拾时依然能够平稳前进。机器可以习得隐性知识，但仍然缺乏人类所拥有的微妙性。

正是这种关键特征在微软引发了关于包容性人工智能的讨论和研究。人工智能领域正在迅速发展，我们都在实验解决方案、应对意想不到的结果的过程中同步学习。作为这一转变的一部分，微软包容设计团队已经与微软的研究、工程和法律团队合作，将高级别原则融入日常实践。

本着知识共享的精神，微软总结了 5 条原则，以期识别排斥，设计更具包容性的人工智能。这些原则在中国业已有了良好实践。

## 1. 重新定义偏见

关于人工智能的讨论，通常是两极分化的“善与恶”，但人工智能产品或服务开发团队很难将自己的实践与那些抓住媒体注意力的、最具冒犯性的偏见案例联系起来。我们了解到，比起关注那些最极端的案例，将偏见看作一个光谱，意识到偏见会以微小的方式显现在日常生活中，会使得团队更快地参与到处理人工智能偏见问题的行动中。

我们要监督人工智能，避免掉入重大陷阱；与此同时，解决微小的、似乎很平常的侵害。想想是否有一个时刻，你觉得某个产品不是为你设计的，但你又无法解释为什么这么觉得。随着时间的推移，这些微小的不舒适逐渐累积，让你产生被排斥感，或者单纯对这个产品不再感兴趣。让人经历“偏见时刻”的产品不是好产品。不过，人工智能偏见并不意味着世界末日，现在只是良好意图出现偏差的早期阶段。我们有责任意识到这些被低估的风险，也有责任进行相应的设计。

我们还可以采取更为主动的行动，而不是坐等问题的发生。一个积极的案例是，2019年世界自闭症日（4月2日）前夕，小米集团联合壹基金推出“为AI发声—蓝色行动计划”。小米与壹基金一起训练“小爱同学”回答与自闭症有关的问题，向公众普及自闭症知识，消除对自闭症群体的刻板印象。

## 2. 招募客户修正偏见

要打造更具包容性的人工智能，训练至关重要。然而，人工智能开发往往是不公开地进行的，因而受制于团队输入的、缺乏多样性的数据，这些数据可能不能代表他们原本计划服务的客户群。团队不仅应考虑他们开发的人工智能产品或服务的应用，还应谦逊地思考当该产品或服务背后的技术被公开使用之后可能产生的影响。赋能用户持续地训练人工智能，将使其更具包容性，并最终建立信任。

过去的数年里，围绕人工智能设计道德的讨论有所增加，议题包括透明度、问责制、代理等。这些概念催生了许多开源项目，比如使用合成数据的自动驾驶模拟平台，以及个人用自己的声音训练语音模型的众包倡议。在安全的训练环境中，尽早理解客户的想法，可以降低出现意料之外或令人担忧的结果的概率。

### 3. 尊重隐私与知情同意

一个普遍的幻想是，随着时间的推移，人工智能将越来越聪明。诚然，机器会从输入的数据中学习，提高理解能力。但包容性人工智能还依赖于数据集本身的多样性、标注的正确性，以及数据以代表每一个用户的方式使用。如果系统中原本就存在偏见（系统中总有偏见），时间只会加剧这种偏见。对于未被充分代表的人们来说，他们毫无动力参与对他们不利的东西，尤其是当他们提供的信息可能被用来针对他们自己。没有这些人的数据，人工智能习得偏见的循环会一直持续。

我们都知道，我们放弃了一些隐私以换取现代生活中的便利。隐私保护方案的设计往往很糟糕，而且对普通用户来说太过复杂。通用数据保护条例（GDPR）的采用促进了行业进步，然而，对隐私的保障应该是基础性的，而不是回应式的。我们需要的不是充斥着难以理解的法律术语的用户协议，而是用户在使用产品或服务的全过程中都有互动途径表达自己的知情同意；我们需要一种将人的自主性置于首位的设计。



## 4. 平衡智能与探索

人工智能基于我们过去的行为做出假设，往往不能灵活地理解我们现在的意图。当这些假设制造出“回音室效应”、“数字气泡”、不相关内容，或者一遍一遍重复的建议时，我们会觉得不舒服。人类总有以全新的方式创造、探索的渴望，这与机器智能之间存在内在张力。我们需要一些关键时机，在人类-技术之间建立更自然的关系，珍视耐心与创造性探索。

客户应该能够始终感觉到，他们有改变方向的选择，有确定自己的体验目标的选择。但我们了解到，情况并非总是如此。我们对可以期待些什么没有明确的认识。我们被引导相信，人工智能是一种“开箱即用”的智能——这样的叙事需要被改变。如果用户知道人工智能服务在起初总是不完美的，是需要他们帮助完善的，他们会更愿意用自身的独特性来训练人工智能。

## 5. 建立包容性团队

尽管我们愿意相信人工智能是中立的，但它确实反映了创造者的想法。雇佣不同背景、学科、性别、种族和文化的人来设计开发至关重要。“人工智能将反映其创造者的价值观，” 纽约大学 AI Now 研究中心的凯特·克劳福德(Kate Crawford)表示，“包容性很重要——从设计者到公司董事会成员，到哪些道德观点被采用，否则，我们创造的，就可能是反映社会狭隘和特权的机器智能，充满熟悉的旧偏见和刻板印象。”

我们知道，拥有多元观点的团队更容易识别偏见。建立包容性团队有利于培养团队成员的同理心，也有利于训练人工智能更具同理心。团队要保持思想开放，对无意的错误负责，以谦虚的态度与公众对话。要考虑周全，深思熟虑，时刻注意自己的设计中是否存在固有的偏见。

## 以人为中心的设计

世界上没有也不应有适用于一切场景的万能公式。如果我们试图打造能真正帮助和理解我们的人工智能，我们就需要从人的角度来看待它。我们不可能仅仅因为自己是未来机器的创造者，就对它们深信不疑——毕竟人类复杂多疑。意识到这一点，或者优雅地失败，都没什么大不了的。在怀疑的时候，我们可以放慢脚步，思考我们为什么要朝某个方向前进，邀请更多的人加入到创作过程中，并不断改进。

人性是有缺陷的，但也很美妙，有着无与伦比的复杂性。我们认为有必要彼此连接，参与解决问题，挑战自己的观点，保持进步。让我们照着最美好的意愿共同行动，通过设计实现更好的人工智能发展。

## 附录

### 联合国可持续发展目标实施背景下《关于科技产品遵守信息无障碍相关原则的北京倡议》

我们，于2017年12月10日国际人权日，为切实落实《中华人民共和国宪法》人权保障的规定，推动联合国《残疾人权利公约》和联合国可持续发展目标在中国的全面履行，在我国“十三五”规划提出一系列明确残障人发展指标并提出同步小康口号的背景下；在听取来自残障人及其亲友、残障人组织对这一议题的见解和思考，了解残障人对其自身在信息时代下权利得以尊重并保障的实际期许，综合考虑残障人权利在中国的现状和挑战后；经过广泛交流、深入研讨形成共识，在北京发出本倡议。

#### 我们明确：

在科学技术高度发展的现代社会，一切有关人类衣食住行的公共空间环境以及各类建筑设施、人机交互设备的规划设计，都必须充分考虑具有不同程度生理机能差异者和活动能力衰退者（如残障人、老年人、儿童等）的使用需求，配备能够应答、满足这些需求的服务功能与装置，以营造一个充满爱与关怀，切实保障人类安全、方便、舒适的现代生活环境；

教育、知识、信息和通信是人类进步、努力和福祉的核心；

信息无障碍对广大残障者有非常重要的意义，在今天甚至比无障碍设施更能改善残障者的生活、学习和工作能力，使他们能够真正平等地参与社会生活；

信息无障碍的发展是针对所有人的，也会让所有人受益；

科技产品在设计生产过程中遵照相关规范进行无障碍设计并非额外负担（若有所缺失，将是产品的

问题与缺陷)；

联合国《残疾人权利公约》的核心精神是：没有我们的参与，请不要做有关我们的决定；

联合国可持续发展目标的口号是：没有一个人会被落下。

### **我们遵守：**

《中华人民共和国宪法》第三十三条、《残疾人权利公约》第九条、《中华人民共和国残疾人保障法》《无障碍环境建设条例》等已有的法律法规。

### **我们倡议：**

所有信息技术及科技产品的研发、设计、生产与销售者，以《残疾人权利公约》为指导，以联合国可持续发展目标为方向，保持开放合作的心态，促进更多参与者的加入；

在产品研发设计、生产与销售的任何一个环节，注重信息无障碍的相关原则；尊重并广泛邀请残障人士在各个环节参与；建立有效的信息无障碍用户体验及意见反馈渠道；积极推动并确立信息无障碍相关规范及培训课程在中国开发者专业发展过程中的重要地位；

使残障人成为《残疾人权利公约》实施与联合国可持续发展目标实现、信息时代社会飞速发展的参与者、推动者和受益者。

### **倡议发起：**

人大法学院-哈佛大学法学院残障项目

一加一残障公益集团

时间：2017年12月10日