arXiv:2304.11565v1 [cs.HC] 23 Apr 2023

专家偏好文本，但视频有助于新手：多媒体内容效用分析

Hayeong Song 佐治亚理工学院 美国亚特兰大
hsong300@gatech.edu

Jennifer Healey
Adobe Research 圣何塞，美国
JeHealey@adobe.com

Alexa Siu
Adobe Research 圣何塞，美国
asiu@adobe.com

Curtis Wigington
Adobe Research 美国圣何塞
wigingto@adobe.com

John Stasko 佐治亚理工学院 美国圣何塞
stasko@cc.gatech.edu

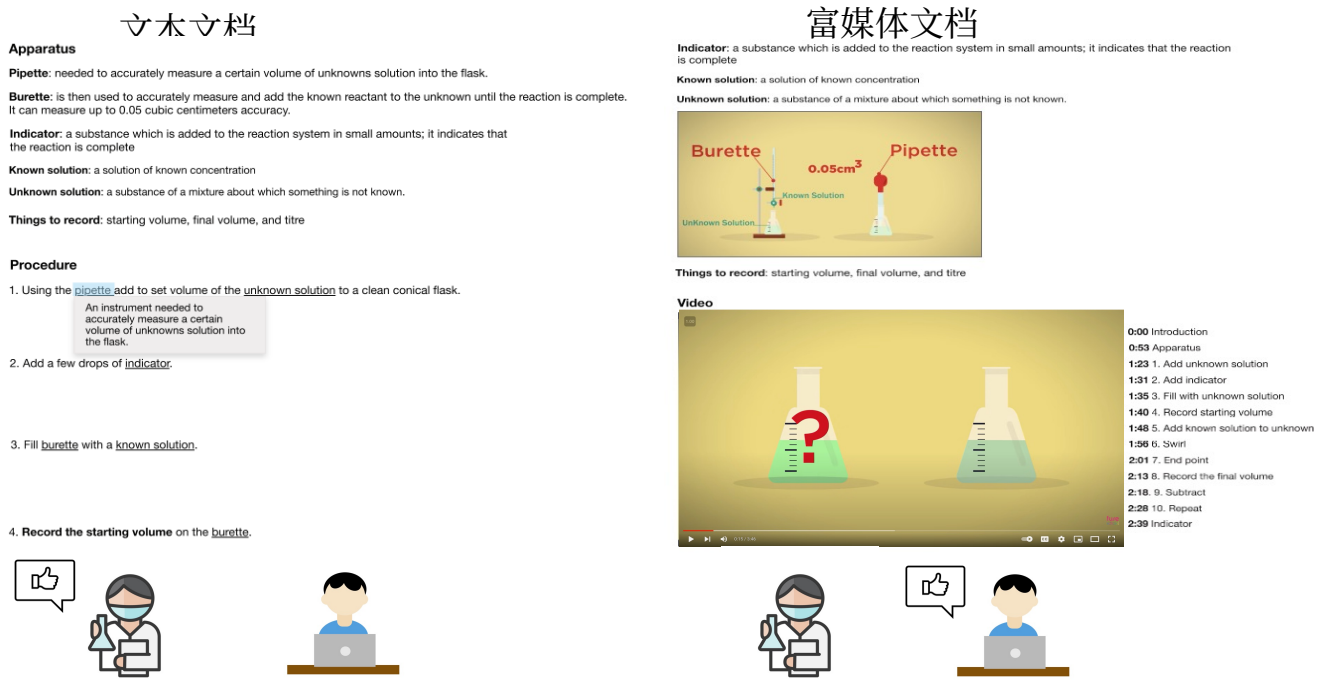

图1：我们为新手和专家设计了两个版本的媒体增强文档。面向专家设计的文本文档（所有信息均为文本）采用了行内文本增强，而面向新手设计的丰富文档（文本、图像、动图和视频）包含多媒体内容。我们在一项双向组间研究中让48名参与者评估了这些原型，结果发现，多媒体内容提升了新手对可用性的感知，并帮助他们缩短阅读时间和测验时间、提高理解程度；而专家基本忽略了多媒体，主要使用文本。

摘要

多媒体能够提高参与度，并且在包括新闻、网络博客和社交媒体在内的在线内容中日益普及，然而，它并不总是对用户有益。为了确定用户实际希望使用哪些类型的媒体，我们开展了一项探索性研究，让用户自行选择媒体增强方式。我们的研究结果表明，用户所期望的媒体数量和类型取决于他们对内容的熟悉程度。为了进一步研究这一差异，我们创建了两个版本的媒体增强文档，一个面向新手设计，另一个面向

特此许可为个人或课堂使用而制作本作品的部分或全部数字或纸质副本，无需支付费用，但前提是不得出于营利或商业利益而复制或分发副本，并且副本必须附有本声明以及首页上的完整引文。本作品中第三方组件的版权必须得到尊重。如需其他用途，请联系所有者/作者。*CHI EA '23，2023*年*4*月*23*日至*28*日，德国汉堡© 2023 版权所有归所有者/作者所有。ACM ISBN 978-1-4503-9422-2/23/04。https://doi.org/10.1145/3544549.3585900

专家设计。我们对这些原型开展了一项包含48名参与者的双因素组间研究，结果发现，尽管多媒体提高了新手读者对可用性的感知（p = .0100），并帮助他们改善阅读时间（p = .0427）、任务用时（p= .0156）和理解（p = .0161），但专家基本忽略了多媒体，主要使用文本。

CCS 概念

• **以人为中心的计算** → **可用性测试**; 人机交互实证研究; • **交互设计** → 用户界面设计；交互设计流程与方法。

关键词

多媒体、文档智能、定制化、新手、专家、增强、阅读条件、文本、图像、动画、GIF、视频

ACM参考格式：

Hayeong Song、Jennifer Healey、Alexa Siu、Curtis Wigington 和 John Stasko。2023。专家偏好文本，但视频有助于新手：多媒体内容效用分析。载于*2023*年计算机系统中人的因素会议（*CHI*）扩展摘要（*CHI EA '23*），*2023*年*4*月*23*日至*28*日，德国汉堡。ACM，纽约州纽约，美国，9页。https: //doi.org/10.1145/3544549.3585900

1 引言

我们认为，在不久的将来，文档将自然地包含交互元素、动画和视频等多媒体内容。我们的这一判断基于以下观察：文本文档持续包含越来越多的图像，而且信息越来越多地通过信息图表、幻灯片和视频演示进行分享。在互联网中，报纸、网络博客和社交媒体等数字文档的多媒体内容也在不断增加 [9]。尽管已有研究表明，多媒体内容能够提高参与度，例如，带有视频的推文分享次数比不带视频的推文多十倍 [40]，，但人们尚未充分研究，多媒体内容对于寻求信息的人来说是否真正所需或有帮助。我们假设，未来的文档将包含多媒体内容，并且数字文档能够针对不同的人，以不同方式定制这些内容的呈现。在本文中，我们介绍了旨在推动此类定制化阅读体验的设计流程。

为了探究人们希望如何增强文本文档，我们与十位参与者开展了一项形成性研究。在研究中，我们向参与者提供了一份关于化学滴定实验的纯文本程序性文档，并请他们在互联网上搜索自认为有助于更好理解文本的多媒体内容。我们选用了一份我们认为可能会给许多人带来挑战的程序性文档，作为本次调查的案例研究。我们注意到，参与者对该主题的熟悉程度不同，搜索次数和检索到的媒体类型也存在差异。由此，我们假设专家会倾向于较少的多媒体内容，而新手会倾向于较多的多媒体内容，尤其是视频内容。因此，我们设计了第二项实验，专门分析新手和专家用户对特定类型文档增强的偏好。

在人机交互社区中，关于增强阅读界面的研究一直在持续，旨在扩展人们的认知并提供流畅的体验，例如根据用户需求调整形式和内容 [12, 21, 34]。然而，较少有研究探讨并比较人们如何消费数字文档中的多媒体功能，以及基于人们对文档内容的熟悉程度（例如新手和专家）的阅读表现。我们的工作采用细粒度的方法，研究新手和专家在多媒体文档消费方面的差异，并将行内文本、图像、动画 GIF 和视频视为增强内容。基于我们的形成性研究，我们提出了以下假设：1）新手将偏好更多的媒体增强，尤其是视频；2）多媒体增强将缩短新手的测验时间，提高其理解程度，并降低其认知负荷。

为检验这些假设，我们针对滴定实验实施了两个版本的数字程序性文档：以文本为重点的版本和媒体丰富的版本。文本版本旨在验证我们关于专家偏好的假设，仅包含纯文本以及针对关键词的行内文本释义。媒体丰富版本旨在验证我们关于新手偏好的假设，其中包含多媒体增强。我们招募了48名参与者，并在嵌套式组间研究中向他们展示这两个原型，以评估定制原型的可用性、记录测验时间，并了解人们消费多媒体功能的模式。我们通过问卷调查和访谈收集数据。结果显示，首先，媒体偏好取决于人们对文档内容的熟悉程度。其次，对于不熟悉该主题的人，多媒体内容能够缩短消费时间并提升理解。最后，对于已经熟悉该主题的人，没有必要加入多媒体内容，但这并不会对该群体的阅读指标产生负面影响。这项工作的主要贡献在于研究和分析人们如何在数字程序性文档中消费多媒体功能，以及评估和分析面向两类不同用户的两种定制化文档体验。

2 相关工作

增强型阅读界面。 在人机交互社区中，阅读界面的基础增强旨在提供“流畅”的体验和定制化体验，以增强人们的认知能力，例如交互式书籍 [21, 34]。这些由人机交互领域引入的原型支持众包答案，以更好地解释网页 [13] 中的内容，或提供预览（例如维基百科预览功能），帮助人们借助导航可供性 [23, 39] 轻松跳转浏览内容，并促进略读。其中一些阅读界面专门致力于通过超文本术语表 [19, 43] 或社交注释 [26]来增强纯文本。例如，这些方法被应用于科学阅读，为临时词 [25] 提供文内定义和局部语境 [38]。这可以减轻他们在互联网上搜索材料的负担，因为搜索结果往往不一致或不相关。这一设计被应用于数字文档阅读场景中的亚马逊 Kindle，在连续行之间的空白处显示棘手词语的释义 [1]。尽管我们的工作受到了这些研究的启发（例如，提供语境化释义 [25]），但这些研究大多聚焦于文本增强，而

较少有研究探讨如何用不同类型的媒体（如图像、动画和视频）增强阅读界面。

新手与专家之间的学习差异。以往研究探讨了新手和专家的学习过程及其不同的学习需求 [11, 18, 20, 37]。人们对主题内容的熟悉程度会影响其参与度 [33, 35] 和理解 [32, 36]。例如，不熟悉内容的人可能会在文档具有连贯流程时学得更好，而熟悉内容的人受到连贯性的影响则较小。由于新手和专家的学习需求不同，这甚至会影响人们的网络搜索行为。例如，熟悉网络搜索且为学科领域专家的读者，在网络搜索中比不具备这些条件的读者更成功 [27, 30]。

为减少搜索的负担，数字文档中出现了纳入多媒体内容的趋势，例如包含视频的报纸。然而，由于新手和专家的学习需求不同，这会影响人们对多媒体内容的理解 [10, 14]。以往研究也探讨了学习者专业水平如何影响多媒体内容的有效性 [15, 28, 29]。这些研究表明，对低知识水平学习者有效的多媒体呈现方式，对知识丰富的学习者则会失去效果（专业知识逆转效应）。若缺乏审慎的设计考量，在未来的文档中加入多媒体内容，实际上可能会损害阅读体验 [42]。然而，人们是否希望看到多媒体内容，以及他们在阅读数字文档时如何使用这些内容，尚未得到充分研究。因此，我们应研究人们如何根据对内容的熟悉程度，以不同方式使用多媒体，并在设计阅读界面以定制化用户体验时将其纳入考量。这类定制化能够使用户受益，并可在用户界面的支持下，帮助新手有效地成长为主题内容方面的专家 [17]。

总而言之，针对大众受众的增强型文档设计，以及新手和专家阅读文本文档时的差异，已经有了大量研究；但研究新手和专家如何以不同方式消费增强型多媒体，尤其是视频的工作相对较少。Leake 等人的工作 [31],启发了我们，他们研究了一般多媒体内容偏好，并开发了一个能够根据文本自动生成幻灯片演示的系统。该工作还超越了一般偏好的研究，探讨了专家用户和新手用户在多媒体消费方面的具体差异。在原型设计方面，与我们的工作最相似的是 ScholarPhi [25]，因为它旨在通过行内文本定义增强阅读界面。我们的工作进一步拓展了这一范围，通过不同丰富程度的媒体内容（图像、动画 GIF 和视频）进一步增强阅读界面。

3 形成性研究

我们想了解人们会如何以及为何选择阅读媒体内容较多或较少的文档。为探究这一问题，我们从便利抽样中招募了十位参与者，开展了一项形成性研究。为了更好地了解用户在媒体消费方面的真实偏好，以及与混合模式文档相关的阅读行为，我们设计了一项实验：向参与者呈现一份文本文档，每个章节后面都设有一处间隔，参与者可以在此搜索更多文本、图像和视频内容，以帮助理解（如图2所示）。我们选取了一份可能会给许多人带来挑战的程序性文档——化学滴定实验——作为本研究的案例研究。

参与者均为研究生，其背景包括研究、开发和设计。参与者通过传单和口碑招募。在开始之前，我们取得了参与者的知情同意，以便录制实验过程。我们将这项研究介绍为一项调查互联网如何帮助他们更好地理解文档的研究。所有参与者都熟悉互联网搜索。会议开始时，为了促使参与者尽力理解材料，我们告知他们最后会有一个必须通过的测验。我们告诉参与者，他们可以根据需要花费足够的时间，并根据需要搜索足够多的材料，以充分理解文档。我们要求参与者在阅读和搜索时“出声思考”。研究期间，我们记录了他们的评论、搜索频率以及他们找到的媒体类型。参与者完成阅读后，我们向他们提供理解测验，并询问一个关于他们对搜索结果满意度的问题。随后，我们进行了半结构式访谈。在访谈中，我们提出了相对开放的启发式问题，重点围绕以下方面：每位参与者使用纯文本文档时的阅读体验、搜索体验，以及他们对不同类型媒体的偏好。

3.1 关键见解

访谈期间，我们询问参与者的阅读体验和搜索体验。参与者谈到文档时，一个显著主题是他们对材料的熟悉程度。我们开始记录参与者的阅读速度、他们搜索的内容类型，以及他们找到所需信息的难易程度；这些似乎取决于他们对文档中的材料原本有多熟悉。随着这一见解逐渐浮现，我们根据参与者的评论（而非任何具体评估），按照我们认为参与者对内容是否“熟悉”来对见解进行编码。比如，如果参与者表示自己是化学专业学生，或表示自己做过多次滴定实验，我们就将其编码为熟悉。而如果参与者表示自己主修文科，并表示以前从未做过滴定实验，我们就将其编码为不熟悉。我们能够明确判断三位参与者熟悉材料（P3、P7、P8），另有三位不熟悉（P1、P6、P10）。对于其余参与者，我们无法作出明确判断（P2、P4、P5、P9），因为他们没有清楚说明自己对内容的熟悉程度。我们将识别出的参与者称为熟悉组和不熟悉组。

在文档上花费的时间。 熟悉的 读者更有可能快速浏览文档，因为他们已经了解了大部分概念。P3，一位快速阅读文档的参与者表示：“先简单介绍一下背景，我本科主修化学。我已经多次进行过滴定实验，因此我对滴定非常熟悉”。 陌生的

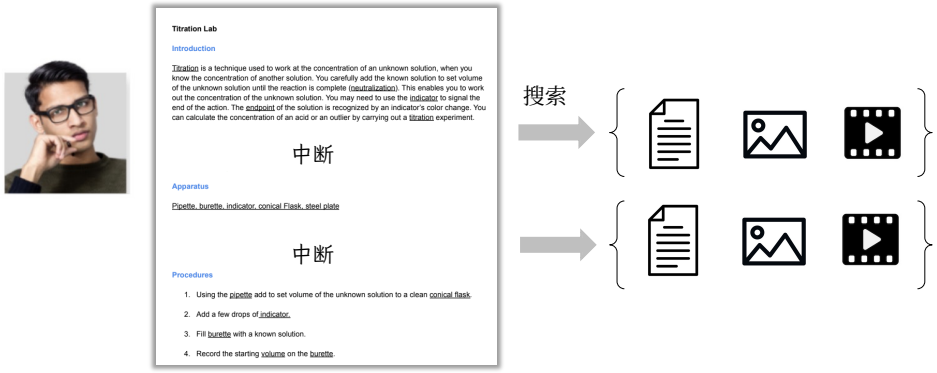

图2：在我们的形成性研究中 在这项研究中，我们开展了一项观察性研究，旨在了解用户对医学 ia 类型（例如，文本、图像、视频 等）的偏好，以及他们在阅读程序性文档时的偏好。

读者在阅读文档和在互联网上搜索陌生术语方面都花费了更多时间，因为他们在阅读过程中逐步了解这些材料。P7表示："我确实不太熟悉化学，所以这将是一篇有趣的文档。*[.....]* 我不知道 *[*这个词*]* 是什么意思，我要查一下。其实，我不知道自己查看的是否是正确的 *[*互联网搜索*]* 结果。 *[.....]* 因为我确实不了解这些 *[*仪器*]*，所以不得不查找它们，这让我花了更长时间阅读。请告诉我，我花的时间是否太多了。"

对互联网搜索结果的满意度。不熟悉该主题的群体对互联网搜索结果的满意度较低，主要是因为他们更难使用正确的搜索词来找到所需内容。P1 表示："我有设计背景。*[.....]* 因为我不熟悉这些术语，所以不知道该查找什么。 *[.....]* 此外，互联网搜索结果令人困惑，因为其中会交替使用不同的术语；如果我不知道某个术语指的是这个还是那个，或者它们是否是同一回事，就会感到困惑。"熟悉该主题的读者对互联网搜索结果的满意度更高，因为他们通常只是试图回忆自己熟悉的特定信息，而不是从头开始学习概念。P8 表示："我本科时学过很多化学，所以已经熟悉这类实验。 *[.....]* 我只需要查几个术语和仪器，回忆它们是什么以及如何使用。 *[.....]* 快速查询公式和搜索定义很有帮助。"

期望的媒体增强。 熟悉的 读者更偏好文本，而不是视频等更丰富的多媒体内容。因为借助文本，他们可以利用自己熟悉的结构快速浏览内容。他们往往不会搜索视频或观看视频，因为他们认为搜索并观看视频需要付出太多努力。P7 表示："我不需要为此观看视频。观看视频会花费太多时间，而且我不想在阅读文档时停下来，或受到打断。阅读更好，或者说文本更好，因为当我不确定时可以查看它。 *[.....]*如果文档中的定义或公式能够得到支持，那就更好了 *[*以文本形式*]*"。相比之下，不熟悉的 读者希望获得并搜索多媒体内容（图像和视频），因为他们想全面了解文档内容的整体图景，并在阅读更多文本信息之前了解更多内容。P10 表示："我的意思是，视频通常总是最有帮助的。因为具体到这个情况，

我最后观看的视频涵盖了我之前查找的所有内容。所以它真的很有帮助。 *[.....]* 更重要的是，视频解释了顺序步骤，这对于准确执行每一步以及理解子步骤也很重要，而这些内容在文本文档中并未涵盖。"

3.2 基于用户偏好的设计考量

本节报告了所有参与者共同提出的用户需求，以及一些针对不太熟悉的读者的特定需求。我们将这些用户需求与设计考量一并说明，并分别称为 D1、D2、D3 和 D4。

- D1. **识别关键词**。10 位参与者中有 6 位表示，他们希望知道哪些词代表了对任务至关重要的项目或概念。当被问及希望如何标示重要词语时，一些参与者建议采用高亮、下划线、粗体、颜色编码或悬停显示信息。P6 表示："因为我不认识这些词，所以不知道该搜索什么。阅读文档时，我不想把所有内容都搜索一遍，因为我对滴定还没有全面的了解。 *[.....]* 最好能标出我需要理解的关键词或重要词语。这样我就可以先尝试理解这些内容。"。

- D2. **提供情境化定义**。所有群体中有 10 名读者中的 6 名表示，希望关键-词语能够提供经过筛选、符合情境的定义（例如滴定中所使用的磁力搅拌器），类似于 [25] 中的做法。这一点对不熟悉组尤其重要，不仅因为这样可以立即获得相关信息，还因为如果缺乏情境，这些参与者就不知道如何编写恰当的搜索词，也无法快速检索到相关内容。P1 表示："如果我能看到情境化搜索结果就好了。我在互联网上搜索内容时，搜索结果过于笼统。我会查看那些图像，但我没有足够的情境来判断自己看到的是否是正确内容。我觉得自己需要一些指导。 *[.....]* 此外，有些结果的解释方式也有所不同

，使用不同术语时也不一样，这让我感到困惑。"来自熟悉组的参与者也希望有行内文本定义，以便快速回忆概念，这样他们就不必费心搜索。P7 表示："阅读时，如果遇到那些我忘记了含义的词，就有点像填空。我会搜索那个词来填补空白。如果文档能提供经过筛选、符合本实验情境的定义就好了，这样我就不用再查找了"。

- D3. **让相关媒体内容易于添加或跳过**。 添加适量的多媒体内容可以带来更好的阅读体验 [41]，让人们保持投入，并有助于概念回忆。我们发现，多媒体内容通常对不熟悉这一主题的群体有所帮助。 所有参与者都会搜索图像和视频，也会查找文本。*P10* 表示："因为我是新手，我对实验和相关内容了解不多，事情很快就变得很专业。尤其是实验步骤方面，视频很有帮助，比如能理解某个步骤会带来什么结果。例如，到了终点，只需滴一滴指示剂，颜色就会立刻改变。能够看到这个视觉目标，有助于理解实验。.....我想在视频和文本之间来回切换：看视频理解实验步骤，再浏览文本，回忆视频里讲了什么。如果视频能分段就好了；我总是用 2 倍速看视频，好快速看完。图像则很适合用来查看特定的仪器"。

他们表示，图像帮助他们了解仪器的外观，而视频则帮助他们更好地理解如何实际进行实验。相比之下，熟悉组几乎只搜索文本信息。尽管他们会查看因文本查询而显示的图像，但不会观看任何视频。熟悉 组表示，他们更偏好文本，因为文本能让他们快速浏览内容。当被具体问及动画 GIF 等替代媒体类型时，所有不熟悉 组成员都表示希望看到展示每个步骤操作方法的动画 GIF，而只有少数熟悉 组成员表示，他们总体上希望看到动画 GIF，但不希望在此内容中使用。P4 表示，" 每个步骤配有动画 *GIF* 会有助于快速了解后续步骤，因为它不会打断阅读，但又有点像每个步骤的快速预览。"

- D4. **支持不同的阅读模式**。 加入多媒体通常会造成干扰，[3, 41], 而且熟悉 组认为它没有价值。由于熟悉 组提到他们会略读文档，我们推测，对于这类人群，排除多媒体内容将减少干扰，并创造更适合略读的阅读体验。

4 为新手和专家设计媒体增强型文档

基于形成性研究的发现，我们希望进一步探究如何专门为熟悉或不熟悉该主题的人设计媒体增强文档。我们现在将熟悉 主题的读者称为专家，将不熟悉 主题的读者称为新手。为此，我们开发了两个不同的原型：一个面向专家的文本 导向文档，以及一个面向新手的媒体丰富 文档，设计依据是第 3.2 节中概述的设计洞见。

4.1 设计流程

我们的原型开发遵循迭代设计流程。为了完善设计，我们分多个轮次反复进行构思、原型设计和获取反馈。在早期原型设计阶段，研究人员每周会面进行构思，随后每隔两到三周开展非正式可用性测试以获取反馈。在这些会议中，研究人员讨论应支持哪些多媒体功能，以及何时、如何在数字文档中支持这些功能。为了获得不同的视角，我们积极招募具有不同专业水平的潜在用户。增强型阅读界面的设计方案通过文献综述（包括超文本 [43], 摘要 [4], 和扩展内容）以及对商业工具的评估（包括 YouTube 的视频时间戳和维基百科的预览功能）得以确定。

4.2 刺激材料文档

我们使用化学滴定实验作为刺激材料文档，以构建具有不同增强形式的多媒体文档。我们将这一场景作为案例研究。考虑到寻找合适的媒体来增强文本文档可能较为困难，化学滴定实验文本取自原本为视频格式的内容。该视频讲解了如何完成一个包含 10 个步骤的滴定实验，包括实验目标、所需仪器和操作步骤。我们将视频转录为纯文本版本。随后，我们保留原始视频，从中提取其他多媒体内容，包括图像、动画 GIF 以及带时间戳的视频。为了使两个原型尽可能具有可比性，我们保持文本完全一致，但为新手 版本添加了不同的媒体类型。我们使用 Adobe XD 和 Anima [5] 插件实现了这些原型。

专家版本。 我们设计了一个以文本为重点的专家版本文档。根据第 3.2 节中概述的设计洞见，我们加入了关键词下划线 和行内文本增强 ，以便于略读和快速回忆关键事实。关键词是在滴定实验中频繁使用的重要词语。我们为这些词语添加了下划线，用户将鼠标悬停在带下划线的词语上时，便可查看文本释义（行内文本）的工具提示。

新手版本。 我们根据设计洞见，为 新手组设计了第二个原型，并加入了 丰富的媒体内容。该版本在 专家版本的基础上增加了 关键词下划线和行内文本释义、描述仪器的关键图像、描述操作步骤的动画 *GIF*，以及带时间戳的书签视频。

5 研究设计

我们想评估定制的数字文档是否为我们确定的每个群体提供了适量的媒体内容：新手 和专家。我们设计了一项任务，让每个群体阅读一份文档，然后参加测验，以了解不同群体如何使用多媒体功能，并衡量他们的理解程度。

我们允许参与者在测验期间查阅文档。我们对这项实验的假设如下：

- **（H1）：** 专家组更偏好文本版本，而新手组更偏好媒体丰富版本。
- **（H2）：** 专家组使用文本版本时阅读速度更快。
- **（H3）：** 新手组阅读丰富版本所花费的时间更长，但完成测验的速度更快、测验得分更高，并且体验到更低的认知负荷。

我们的研究设计针对文档的每个版本分别开展双因素被试间研究。我们通过筛选问卷招募了独立的专家和新手队列，让他们体验媒体丰富文档或文本文档。我们总共招募了48名参与者（24名专家和24名新手）。在每个队列中，参与者被随机分配到文档的文本版本或媒体丰富版本。

参与者。我们从不同背景的人群中招募了48名参与者（女性：22名，男性：26名），其背景包括工程学、教育、化学、产品管理和设计。对于主持式会话，研究耗时40分钟完成，我们向参与者提供了一张35美元的礼品卡作为补偿。对于非主持式会话，研究耗时20分钟，我们向参与者提供了一张10美元的礼品卡作为补偿。主持式会话与非主持式会话完全相同，唯一不同之处在于：主持式会话的说明由主持人口头给出，会话结束后还会进行半结构化访谈。在非主持式版本中，参与者阅读说明。主持式会话的用户通过口头宣传和传单招募，非主持式会话的用户则通过User Testing [2] 平台招募。为识别新手和专家，我们使用筛选问卷询问潜在参与者的背景和专业领域。要符合专家资格，参与者需要正确回答这些问题中的大多数（专家正确率高于70%，新手低于50%，共10道题）。

流程。对于符合条件的参与者，我们会征求其同意参与研究。我们告知所有参与者，他们可以根据需要花任意多的时间阅读文档，以充分理解其内容。开始时，我们告诉他们阅读结束后要参加理解测验，但测验期间可以查阅文档。随后，参与者可以自然地阅读文档，不再接受进一步指导（无指导阅读）。每个原型的完整版本如图1所示。每位参与者读完后都会参加理解测验，接着填写调查问卷，以收集可用性（系统可用性量表） [8] 和认知负荷 [24]的量化测量结果。大多数实验采用非主持形式，通过远程测试平台（User Testing [2]）进行。我们从每种条件中随机选出两位参与者，共八位，参加主持形式的研究。

6 结果

在本节中，我们报告了客观绩效测量结果，包括阅读时间、测验时间和理解得分；也报告了主观测量结果，包括可用性、NASA任务负荷指数中的心理需求和努力程度得分。这些结果汇总于表1–4，并在下文详细说明。此外，我们还纳入了主持式会话中半结构化访谈的定性反馈。

6.1 客观测量。

我们从录制的屏幕会话视频中记录了阅读时间和测验完成时间。我们还记录了参与者在媒体内容丰富的文档条件下是否观看了视频。t检验结果汇总于表1和表2。

阅读时间。我们发现，新手在媒体丰富（均值 = 319.58秒，标准差 = 101.08）文档与文本（均值 = 213.58秒，标准差 = 137.54）文档之间的阅读时间存在显著差异（t(22) = 2.15，p = .04）。但在不同条件下，专家的阅读时间没有显著差异（t(22) = 1.81，p = .08）。这意味着，与文本条件下的新手相比，媒体丰富条件下的新手花费的时间显著更长，并且使用了文档中的多媒体功能。

我们注意到，在多媒体丰富的文档中，新手更常选择从头到尾观看视频（12 名参与者中有 10 名），而且所有新手都以正常速度播放视频。专家中只有一半（12 名参与者中有 6 名）选择观看多媒体丰富文档中的视频，且大多数专家只观看了视频的一部分。

测验时间。我们同样发现，在多媒体丰富的文档（M = 149.99s，SD = 45.92）和文本文档（M = 299.15s，SD = 191.82）中，新手的测验时间存在显著差异（t(22) = -2.61，p=.01）。但我们未发现专家在不同条件下存在显著差异（t(22) = 0.51，p = .61）。这意味着新手受益于多媒体丰富的文档，因为他们完成测验所用的时间比文本条件下的新手更短。另一方面，专家并未从多媒体内容中显著受益。

我们还发现，在完成文本文档的测验方面，专家 和新手之间存在显著差异：文本 文档条件下，专家 花费的时间为（均值 = 164.97 秒，标准差 = 92.02），而新手 花费的时间为（均值 = 299 15 = 191 82 22 = 2 18, = 03 . 秒，标准差 . ），（t( ) - . p . ）。这意味着，文本条件下的新手 完成测验所花费的时间显著长于文本条件下的专家 。

测验得分。我们再次发现，在文本文档（M = 6.91，标准差 = 2.02）与媒体丰富文档（M = 8.66，标准差 = 1.15）之间，新手的表现存在显著差异（t(22) = 2.60，p = .01）。但我们未发现专家在不同条件间存在显著差异（t(22) = 0 42, = 67 . p . ）。这意味着，在媒体丰富条件下，能够访问多媒体内容的新手测验得分显著高于文本中新手条件下的参与者。在两种条件下，无论是否能够访问多媒体内容，专家的测验得分都很高。我们推测，这是因为专家已经熟悉大部分内容，并且他们使用程序性文档来回忆概念。

我们还发现，专家与新手在文本文档的测验分数方面存在显著差异：专家的得分为（M = 8.58，SD = 0.66），而新手的得分为（M = 6.91，SD = 2.02），（t(22) = 2.71，p = .01）。这意味着，文本条件下的专家得分显著高于文本条件下的新手。

6.2 主观测量：可用性与认知负荷。

在我们的主观分析中，我们要求参与者完成一份调查问卷，其中包括用于评估可用性的SUS量表问题 [8]，以及用于评估感知心理需求和努力程度的NASA-TLX问题 [24]（见表3和表4）。

表1：客观测量

条件	阅读时间（秒）	测验时间（秒）	测验得分
新手 ,rich	均值 = 319.58 , 标准差 = 101.08	均值 = 149.99 , 标准差 = 45.92	均值 = 8.66 , 标准差 = 1.15
新手 , text	均值 = 213.58 , 标准差 = 137.54	均值 = 299.15 , 标准差 = 191.82	均值 = 6.910 , 标准差 = 2.02
专家，丰富	均值 = 251.41，标准差 = 121.86	均值 = 187.53，标准差 = 118.85	均值 = 8.83，标准差 = 0.71
专家，文本	均值 = 175.41，标准差 = 79. 36	均值 = 164.97，标准差 = 92.02	均值 = 8.58，标准差 = 0. 66

表2：T检验客观指标分析。* 表示显著性

比较	阅读时间	测验时间	测验得分
新手，丰富 & 新手，文本	***t(22) = 2.15，p = .04**	***t(22) = -2.61，p = .01**	***t(22) = 2.60，p = .01**
专家，丰富 & 专家，文本	t(22) = 1.81，p =.08	t(22) = 0.51，p =.61	t(22) = 0.42，p =.67
专家，文本 & 新手，文本	t(22) =-0.83，p = .41	***t(22) = -2.18，p = .03**	***t(22) = 2.71，p = .01**
新手，丰富 & 专家，丰富	t(22) = 1.49，p =.15	t(22) = -1.02，p =.31	t(22) = -0.42，p =.67

可用性。根据SUS 分数，我们发现新手 在可用性得分方面存在显著差异: 文本 文档（均值 = 63.95（低于平均水平），标准差 = 18.56）与媒体丰富 文档（均值 83 33 = 14 93 22 = 2 81， = 01 .（优秀），标准差 .）之间存在显著差异，（t( ) . p . ）。但是，我们没有发现专家 在不同条件下存在显著差异，（t(22) = 1.52，p = .14）。这意味着，与文本条件下的新手 相比，媒体丰富 条件下的新手 报告了显著更高的可用性得分。

但我们没有发现，多媒体专家和文本专家条件之间的感知可用性分数存在显著差异，因为他们基本上忽略了多媒体功能，并且对行内文本增强感到满意。我们还发现，文本专家条件的报告可用性得分（均值 = 75.83（良好），标准差 = 16.69）高于文本新手条件（均值 = 63.95（低于平均水平），标准差 18 56 22 =2 20， = 03 .），（t( ) . p . ）。

心理需求与工作负荷。我们未发现参与者在各条件下的感知心理需求和努力程度存在显著差异。

6.3 定性反馈

我们收集了参加主持式研究的八名参与者的定性反馈。这包括来自四个分配条件中每个条件的两名参与者，分别称为 新手、丰富 NR、新手、文本 NT、专家、丰富 ER 和 专家、文本 ET。

新手。与文本 版本互动的(NT1-2) 新手希望有更丰富的媒体内容，例如图像和视频，以说明仪器的外观。NT2 表示 "除了文本之外 *[*，我还希望有图像或视频*]*，更详细地说明每个步骤。 *[...]* 我觉得区分实验中使用的仪器有点令人困惑。不过，我发现工具提示的定义很有帮助，而且易于使用。"与丰富版本互动的新手发现多媒体内容有助于理解文档内容（NR1-2）。NR1 说 "我发现这些视频有助于了解如何进行滴定实验的步骤。如果你注意到了，我在做测验时会回看动画 *GIF*，以回忆一些需要澄清的内容。"NR2 表示 "由于我不了解这些材料，通过视觉方式查看这些材料很有帮助。尤其是，我喜欢视频和动画 *GIF* 的这一点。解释我正在阅读的材料时一字不差。 *[.....]*阅读时，如果我看不懂了，就可以去看视频或动画 *GIF*，继续读下去。预先提供这些信息的 *[*可视化呈现*]* 真好，这样我就不用自己去查找这些材料了。"

专家。与文本版本互动的专家表示，他们掌握了所需的全部信息，并不希望增加额外的媒体内容（ET1-2）。他们还认可行内文本增强。ET1表示 "我主修工业化学，所以我知道如何进行滴定实验，而且我已经做过多次。 *[...]* 这份文档对我来说已经足够了，其中包含了我需要的全部信息。 *[...]* 我不想观看视频之类的内容"。ET2表示，"我熟悉滴定，但有时需要回忆一些概念。滴定实验有不同类型，例如酸碱滴定以及其他类型。 *[...]* 我喜欢那个工具提示，其中的定义能立即解释我正在查找的内容，这正是我需要的。而且这些解释与上下文直接相关， *[*这些解释是针对具体实验类型定制的*]*"。与该丰富文档互动的专家似乎没有关注多媒体内容（ER1-2）。ER1表示 "我确实想过观看视频，因为视频就在那里。但我意识到文本中包含相同的信息，而阅读文本更容易。我可能看过动画 *GIF*，因为它们就在那里，但我主要还是阅读文本，并在阅读文档时遇到困难便使用工具提示。 *[...]* 如果要观看视频，我必须寻找一个 *[*视频片段*]* 来解释相关内容，而这太麻烦了。另外，如果我开始观看视频，就会打断阅读，你知道的，而我不想停止阅读"。据观察，两位专家在主持式会话中与媒体丰富版本互动时，似乎只是略读文档。他们只使用了少量行内文本弹窗来回忆概念，在主持式会话中，双方专家都没有观看视频的任何部分。

7 讨论

我们希望开发一份程序性文档，为所有用户提供适量的媒体内容。为此，我们开展了一项形成性研究，以了解人们如何阅读纯文本

表3：主观测量

条件	可用性得分	心理需求	努力程度
新手，丰富	均值 = 83.33，标准差 = 14.93	均值 = 3.5，标准差 = 1.80	均值 = 3.5，标准差 = 1.83
新手，文本	均值 = 63.95，标准差 = 18.56	均值 = 4.27，标准差 = 2.24	均值 = 3.91，标准差 = 2.02
专家，丰富	均值 = 87.5，标准差 = 15.73	均值 = 3.18，标准差 = 1.40	均值 = 2.83，标准差 = 1.11
专家，文本	均值 = 75.83，标准差 = 16.69	均值 = 3.09，标准差 = 1.57	均值 = 2.7，标准差 = 0.94

表4：t检验主观测量分析。* 表示显著性

比较	可用性得分	心理需求	努力程度
新手，丰富 & 新手，文本	***t(22) = 2.81，p = .01**	t(22) = -0.83，p = .41	t(22) = -0.52，p = .60
专家、丰富 & 专家、文本	t(22) = 1.52，p =.14	t(22) = 0.11，p =.91	t(22) = -0.61，p = .54
专家、文本 & 新手、文本	***t(22) = 2.20，p = .03**	(22) = -1.01，p =.32	t(22) = -1.02，p =.31
新手、丰富 & 专家、丰富	t(22) = -0.66，p = .51	t(22) = 0.11，p =.90	t(22) = 1.07，p =.29

程序性文档，并调查他们还希望获得哪些类型的媒体内容。我们的分析表明，人们对主题内容的熟悉程度会影响他们阅读文档和使用多媒体功能的方式。这促使我们设计了第二项实验，专门针对新手和专家用户。我们实现了媒体增强文档的两个版本：文本文档和媒体丰富文档。在第二项实验中，我们评估这两种文档是否为我们所确定的群体提供了适量的媒体内容：新手和专家。结果表明：

- H1：（正确）专家组将偏好文本版本，而新手组将偏好媒体丰富型版本。总体而言，专家更偏好文本（增强版），因为该版本满足了他们所有的信息需求。他们基本忽略了媒体丰富型版本中的多媒体内容，这一点从主持式访谈的定性结果中可以看出。
- H2：（不正确）专家组将使用文本版本阅读得更快，并更快完成测验。
- H3：（基本正确）新手组阅读媒体丰富型版本所花费的时间会更长，但他们将更快完成测验，在测验中取得更高分数，并体验到更低的认知负荷。

用户对多媒体消费的偏好。我们的结果表明，多媒体内容，尤其是视频，并非始终受到用户青睐，而是取决于用户以及他们对文档内容的熟悉程度。我们认为，定制文档中包含的多媒体内容数量，包括图像、动图和视频，能够提高生产力并减少干扰。我们设想，可以通过多种方式实现可变的多媒体插入。一种简单的方法是提供一个用户界面元素，让用户根据自身偏好手动选择媒体内容。另一种方法是根据用户的阅读历史自动识别新手和专家 [33]。鉴于许多网站和电子阅读器都会收集有关用户行为的信息，我们假设可以根据这些收集到的数据推断用户对特定主题的熟悉程度，随后网站便可以提供插件功能，以支持适当丰富度级别的多媒体内容，例如加入更多视频或仅提供增强文本。我们设想，我们的工作可以应用于其他情境和场景，例如根据学习者专业水平定制多媒体支持程度，应用于教育 [14],，如化学、医学、词汇学习 [6, 7, 16]，以及文化遗产旅游中面向专家和新手访客的场景 [10]。

动态多媒体增强。我们还认为，可以在读者阅读文档的过程中，利用预测统计模型 [22, 45]，根据阅读速度或其他行为动态评估用户对内容的熟悉程度。在我们的研究中，我们呈现了两份固定内容的文档，一份包含多媒体内容，另一份不包含多媒体内容；然而，也可以根据实时用户行为动态加入这些内容。例如，当阅读速度变慢时，文档可以将文本转换为图像 [44],，将文本转换为视频，以更好地解释内容 [44],；算法甚至可以将整份文档转换为视听幻灯片演示 [31]。动态改变多媒体内容的数量支持自动定制，使文档能够适应不断变化的用户偏好，例如始终能够添加内容（扩展信息）。

8 结论

我们的高层目标是为所有用户开发媒体增强程度恰到好处的数字文档。为此，我们对真实用户阅读一份程序性文档进行了观察性研究。研究发现，人们会根据自己对内容的熟悉程度，以不同方式使用数字文档中的多媒体功能。为验证我们的假设——人们对多媒体内容的需求取决于他们对材料的熟悉程度——我们随后开展了第二项用户研究，邀请了48位不同的参与者；我们特意招募了化学领域的新手和专家，让他们阅读一份以文本为主或媒体内容丰富的化学滴定实验文档。研究结果表明，专家更喜欢文本，而新手则能从多媒体内容中受益。也就是说，针对不同类型的用户定制文档中的多媒体丰富程度，能够提高生产力并减少干扰。

参考文献

[1] 2020。阅读时查找词语、人物和地点。https://www.amazon. com/b?ie=UTF8&node=17717476011[2] 2023。User Testing。https://www.usertesting.com/platform?utm_source=mkto- fallback[3] Jesper Aagaard。2019。多任务处理作为干扰：媒体多任务处理研究的概念分析。Theory & Psychology 29, 1 (2019), 87–99。[4] Mehdi Allahyari、Seyedamin Pouriyeh、Mehdi Assefi、Saeid Safaei、Elizabeth D Trippe、Juan B Gutierrez 和 Krys Kochut。2017。文本摘要技术：简要调查。arXiv 预印本 arXiv:1707.02268 (2017)。[5] Anima 2022。适用于 Adobe XD 的 Anima。https://docs.animaapp.com/v3/adobe- xd/getting-started/。[6] Henriette Arndt 和 Robert Woore。2018。通过观看 YouTube 视频和阅读博客文章学习词汇。Language Learning and Technology 22, 3 (2018)。[7] Thomas Balslev、Willem S De Grave、Arno MM Muijtjens 和 AJJA Scherpbier。2005。在研究生问题导向学习课程中比较文本案例和视频案例。Medical education 39, 11 (2005), 1086–1092。[8] John Brooke。1996。“SUS——一种快速简便的可用性量表。”《工业领域的可用性评估》。CRC出版社。https://www.crcpress.com/product/isbn/9780748404605 ISBN: 9780748404605。[9] Kyle Byers。2022。2022年151项内容营销统计数据。https://growthbadger. com/content-marketing-statistics/ [10] Heloisa Candello。2013。移动文化应用中的多媒体信息传递。载于《国际人机交互普适访问会议》。Springer，583–592。[11] Anne Castles、Kathleen Rastle 和 Kate Nation。2018。终结阅读之争：从新手到专家的阅读习得。Psychological Science in the Public Interest 19, 1 (2018), 5–51。[12] Bay-Wei Chang、Jock D Mackinlay、Polle T Zellweger 和 Takeo Igarashi。1998。一种流式文档协商架构。载于第11届年度ACM用户界面软件与技术研讨会论文集。123–132。[13] Parmit K Chilana、Amy J Ko 和 Jacob O Wobbrock。2012。LemonAid：基于选择的众包情境帮助，适用于网络应用。载于SIGCHI计算机系统人为因素会议论文集。1549–1558。[14] Thomas KF Chiu、Morris Siu-yung Jong 和 Ida AC Mok。2020。设计面向小学生数学学习者的情感多媒体时，学习者的专业水平是否重要？Educational Technology Research and Development 68, 5 (2020), 2305–2320。[15] Thomas KF Chiu 和 Ida AC Mok。2017。多媒体学习中的学习者专业水平与数学高阶思维技能。Computers & Education 107 (2017), 147–164。[16] Veronica Margherita Cocco、Elisa Bisagno、Gian Antonio Di Bernardo、Alessia Cadamuro、Sara Debora Riboldi、Eleonora Crapolicchio、Elena Trifiletti、Sofia Stathi 和 Loris Vezzali。2021。比较故事阅读和视频观看这两种不同形式的替代性接触：一项针对小学生的实验性干预。British Journal of Social Psychology 60, 1 (2021), 74–94。[17] Andy Cockburn、Carl Gutwin、Joey Scarr 和 Sylvain Malacria。2014。支持用户界面中从新手到专家的过渡。ACM Computing Surveys (CSUR) 47, 2 (2014), 1–36。[18] Harry M Collins 和 Robert Evans。2002。科学研究的第三次浪潮：关于专业知识与经验的研究。Social studies of science 32, 2 (2002), 235–296。[19] Jeff Conklin。1987。超文本：简介与调查。computer 20, 09 (1987), 17–41。[20] 美国国家研究委员会等。2000。《人们如何学习：大脑、心智、经验与学校：增订版》。美国国家科学院出版社。[21] Dennis E Egan、Joel R Remde、Louis M Gomez、Thomas K Landauer、Jennifer Eber-hardt 和 Carol C Lochbaum。1989。Superbook的形成性设计评估。ACM Transactions on information systems (TOIS) 7, 1 (1989), 30–57。[22] Enrique Frias-Martinez、Sherry Y Chen 和 Xiaohui Liu。2006。面向自适应超媒体用户建模的数据挖掘方法调查。IEEE系统、人及控制论汇刊C辑（应用与评述）36, 6 (2006), 734–749。[23] Jamey Graham。1999。读者助手：个性化文档阅读环境。载于SIGCHI计算机系统中的人为因素会议论文集。481–488。[24] Sandra G Hart。2006。NASA任务负荷指数（NASA-TLX）；20年后。载于人为因素与工效学学会年会论文集，第50卷。Sage出版社，Sage CA：洛杉矶，加利福尼亚州，904–908。[25] Andrew Head、Kyle Lo、Dongyeop Kang、Raymond Fok、Sam Skjonsberg、Daniel S Weld 和 Marti A Hearst。2021。为科学论文添加即时、位置敏感的术语和符号定义。载于2021年CHI计算机系统人为因素会议论文集。1–18。[26] William C Hill、James D Hollan、Dave Wroblewski 和 Tim McCandless。1992。编辑痕迹与阅读痕迹。载于SIGCHI计算机系统中的人为因素会议论文集。3–9。[27] Christoph Hölscher 和 Gerhard Strube。2000。互联网专家和新手的网络搜索行为。计算机网络33，1-6（2000），337–346。[28] Slava Kalyuga。2013。学习者先验知识和工作记忆限制对多媒体学习的影响。*Procedia-SocialandBehavioral Sciences* 83（2013），25–29。[29] Slava Kalyuga。2014。多媒体学习中的专业知识逆转原则。（2014）。[30] Ard W Lazonder、Harm JA Biemans 和 Iwan GJH Wopereis。2000。新手与有经验的用户在万维网上搜索信息时的差异。美国信息科学学会期刊51，6（2000），576–581。[31] Mackenzie Leake、Hijung Valentina Shin、Joy O Kim 和 Maneesh Agrawala。2020。使用词语具体性从文本文章生成视听幻灯片演示。载于*2020*年*CHI*计算机系统中的人为因素会议论文集。1–11。[32] Danielle S McNamara 和 Walter Kintsch。1996。从文本中学习：先验知识和文本连贯性的影响。话语过程22，3（1996），247–288。[33] Sungjin Nam、Zoya Bylinskii、Christopher Tensmeyer、Curtis Wigington、Rajiv Jain 和 Tong Sun。2020。使用来自移动设备的行为交互对读者先前熟悉度和目标条件进行分类。*arXiv*预印本 *arXiv:2004.12016*（2020）。[34] Don Norman。2013。日常事物设计：修订扩展版。Basic books。[35] Heather L O'Brien、Luanne Freund 和 Richard Kopak。2016。探究用户参与度在数字阅读环境中的作用。载于*2016*年*ACM*人类信息交互与检索会议论文集。71–80。[36] Yasuhiro Ozuru、Kyle Dempsey 和 Danielle S McNamara。2009。先验知识、阅读技能和文本衔接对科学文本理解的影响。学习与教学19，3（2009），228–242。[37] Charmi Patel。2018。*Jean Lave* 和 *Etienne Wenger* 的《情境学习：合法的边缘性参与》分析。Macat Library。[38] Napol Rachatasumrit、Jonathan Bragg、Amy X Zhang 和 Daniel S Weld。2022。CiteRead：将局部化引文语境整合到科学论文阅读中。载于第*27*届智能用户界面国际会议。707–719。[39] Bill N Schilit、Gene Golovchinsky 和 Morgan N Price。1998。超越纸张：使用自由形式数字墨水注释支持主动阅读。载于*SIGCHI*计算机系统人为因素会议论文集。249–256。[40] Matt G. Southern。2019。*Twitter*揭示带视频的推文可产生高出*10*倍的参与度。https://www.searchenginejournal.com/twitter-reveals-tweets-with-video-generate-10x-more-engagement/322523/#close [41] Zsofia K Takacs、Elise K Swart 和 Adriana G Bus。2015。技术增强型故事书中多媒体和交互功能的益处与陷阱：一项元分析。教育研究评论85，4（2015），698–739。[42] Barbara Tversky、Julie Bauer Morrison 和 Mireille Betrancourt。2002。动画：它能促进理解吗？国际人机研究期刊57，4（2002），247–262。[43] Polle T Zellweger、Bay-Wei Chang 和 Jock D Mackinlay。1998。用于支持信息获取和渐进式链接过渡的流动链接。载于第九届*ACM*超文本与超媒体会议论文集：链接、对象、时间与空间——超媒体系统中的结构：链接、对象、时间与空间——超媒体系统中的结构。50–57。[44] Xiaojin Zhu、Andrew B Goldberg、Mohamed Eldawy、Charles R Dyer 和 Bradley Strock。2007。用于增强交流的文本到图片合成系统。载于*AAAI*，第7卷。1590–1595。[45] Ingrid Zukerman 和 David W Albrecht。2001。用于用户建模的预测统计模型。用户建模与用户适应性交互11，1（2001），5–18。