尊敬的评委老师们：

大家晚上好！

我是“明甄谛听”项目的负责人蒋青江，很荣幸能够在这里向大家汇报我们基于对抗式网络的AIGC检测系统的研究成果。

在当前这个信息爆炸的时代，AIGC技术的广泛应用为我们带来了便捷，但同时也带来了信息真伪难辨的挑战。因此，我们团队致力于研发一款高效、准确的AIGC检测系统，为信息安全保驾护航。

接下来我将从政策导向、行业现状、产品介绍、团队介绍以及前景规划方面讲解，向大家详细介绍。

首先是政策导向这一部分，在图片中我们可以看到国家政府对生成式人工智能管理的高度重视，出台了生成式人工智能服务的管理暂行办法，政策的出台说明了**促进**生成式人工智能健康发展和**防范**生成式人工智能服务风险是当前现实需要。

接下来是我们的第二部分行业现状，首先是生成式AI在艺术创作领域中的影响，在艺术创作领域中生成式AI可能会导致创意的丧失和引发知识产权的争议，在右图中我们可以看到一些知识产权纠纷的一些事例。然后是第二点生成式AI在信息领域的影响，生成式AI在信息领域可能导致用户隐私泄露，也可能通过大数据“杀熟”使部分群体受到算法歧视，由于生成式AI能够生成逼真的文本和媒体内容，就可能会产生谣言和信息误导。第三点，生成式AI在学术领域的影响，在学术领域可能会造成学术诚信问题，信息缺乏可信度、学术失业风险，右侧是一些使用AI的学术不端的问题事例。最后讲解一下AIGC检测平台的行业现状，在国际方面目前已存AI生成作品检测系统**发展缓慢**，其中基本**只包含了文案检测功能**，检测**准确率低，目前市面上已有的检测技术包括有：知网学术文献检测系统（已添加AI文本检测功能）、小型检测网站（几乎为国外网站）等，对中文检测效果更差。在国内方面 国内目前存在的AIGC检测系统仅仅针对学术文本进行检测，缺乏其他如图像、音频、视频的检测系统，国内检测网站收费也普遍较高，不适用于中低端市场（大量高校作业论文检测与中学作业检测）。最后是**我们项目的技术难点，主要有以下四点：1. AI生成作品**与人类作品具有高度相似性**，计算机难于区分。2. 生成式**AI的发展极为迅速**，这也使得检测技术需要有高度成长性。3. **传统文本查重与图像相似度检测**等技术**不适用**于AI生成类作品的检测。4. **没有完善的管控手段**。与系统的处理方案，导致如今AI技术越发难于管理

接下来进入产品介绍部分，首先是产品演示，右侧是AI生成的一段见解类文案，左侧展示的是“明甄谛听”对该文案的检测结果，目前对于单段文案，我们已经取得了不错的效果。

然后是技术介绍，项目采用两项重要技术，第一项技术是**成长式模型训练技术**

在项目开发过程中，我们采用了对抗式网络的先进技术，**对抗式技术：**以生成对抗网络模型为基础核心，该网络模型一般是一个生成模型与判别模型之间进行对抗与训练得出的，当生成模型的仿真能力超过了判别模型就可以得到了我们所知的AIGC模型。而我们的技术便是通过团队内部技术成员自行设计的训练算法与对抗逻辑，训练出一个可以应对生成模型的判别模型。目前得出的模型便是Beta版预测模型——“谛听DIS-1.0”。

总的来说就是AIGC采用生成模型，而我们自己训练出一个可以应对生成模型的判别模型。

第二项技术是**隐私保护训练技术，联邦学习技术：联邦学习技术可以在客户端就进行训练与参数返回，不用大规模采集用户数据，极大的保护了用户的数据隐私。**

接下来是我们产品的核心竞争力，主要有以下四点：第一点，具有强大的市场潜力，目前的**国家政策与国内市场需求**，对于AIGC的管控需要进一步加强。第二点，有足够的创新，具有**区别于传统**查重系统与目前国内AI检测系统，我们开发了**图像检测与音视频检测。**第三点，保障数据隐私，采用联邦学习技术的使用可以极大的保护用户数据安全，**防止数据泄露**。

最后一点，**面向中低端技术市场**，我们模型部署完成后的检测成本极低

最后是我们的目标客户，“明甄谛听”——基于联邦对抗式网络的AI参与度检测系统

我们的产品是一款可以对AI参与工作程度进行检测的多平台检测系统，可以在网页端、PC应用程序、移动端应用进行使用，类似使用案例有用于如哔哩哔哩、抖音、腾讯视频等音视频平台可以部署进行作品的初步审查，判断其AI参与制作的程度；也可以与各类学习软件，例如学习通、中国大学慕课等多类学习软件进行后台部署，便以其进行作业AI参与度检测；或是可以用于小型报刊、小型网络媒体新闻资讯审查等。产品的核心理念是为中低端文献以及日常创作等常见领域进行初步检测，其检测对象则为AI参与程度，为未来规范化管理生成式人工智能做出些许贡献。

下一部分我们介绍一下我们的团队，**我们的团队始于三位对人工智能充满热情的成员。随着项目的不断发展，我们逐渐增加了新的队员，直至现在的团队。每个新成员都带来了不同领域的专业知识和技能，这使得我们的项目不断得到完善和扩展。**

最后一个部分前景规划，目前的系统**正处于开发阶段初期**，暂时不具有大规模应用的能力，我们正在对模型进行完善，对算法技术进行更新，**预期2024年中将具有完备的部署能力**。

彼时，我们将寻找初始合作伙伴进行**首次系统商用测试**，然后进一步实现我们的后期规划。AIGC作品的管控方案亟待完事，在此背景下，设计与开发一款检测AIGC的系统是未来AI规范使用的必行之路，我们愿为此展现出自己的一份力量。

紧接着是我们项目未来的目标，主要有以下四点，第一点资金投入技术研发，提升产品技术这个核心竞争力，第二点与小型企业达成试用协议帮助检测系统的进一步完善，第三点开拓更加广阔的市场，进一步提升产品综合素质，第四点产品占据更大市场，得到更大发展空间与资金。

然后是我们项目的销售模式，有以下五点，1.**网页销售模式2.许可证销售模式3.基于订阅服务的销售模式4.定制化解决方案模式5.授权模式**

**最后我们的团队正致力于开发基于人工智能和大数据的技术，某种程度上也是未来人工智能技术革命的缩影。即使我们的项目可能会遭遇失败，也为实现未来全国的人工智能技术革命贡献自己的一份力量！**

“明甄谛听“筑好AI时代的最后一道防线，以上就是我们项目的全部内容。

谢谢大家！

尊敬的评委、老师们：

大家晚上好！

我是“明甄谛听”项目的负责人蒋青江，很荣幸能够在这里向大家汇报我们基于对抗式网络的AIGC检测系统的研究成果。

在当前这个信息爆炸的时代，AIGC技术的广泛应用为我们带来了便捷，但同时也带来了信息真伪难辨的挑战。因此，我们团队致力于研发一款高效、准确的AIGC检测系统，为信息安全保驾护航。

接下来我将从政策导向、行业现状、产品介绍、团队介绍以及前景规划方面讲解，向大家详细介绍。

首先是政策导向这一部分。在国家政府高度重视生成式人工智能管理的背景下，我们参照生成式人工智能服务管理暂行办法，旨在促进生成式人工智能健康发展与风险防范。

接下来是我们的第二部分行业现状。生成式AI在艺术创作领域可能引发创意丧失和知识产权争议，在信息领域可能导致用户隐私泄露以及大数据“杀熟”问题。而目前国内外的AIGC检测系统仍存在技术不足、检测范围有限等问题。

接下来进入产品介绍部分。我们的AIGC检测系统利用对抗式网络技术，采用成长式模型训练和隐私保护训练技术，具备图像、音频、视频等多维度检测能力。我们的核心竞争力在于市场潜力、创新性、数据隐私保护和低成本检测。

接下来是团队介绍。我们团队由对人工智能充满热情的成员组成，具备多领域专业知识和技能，致力于不断完善和拓展我们的项目。

最后是前景规划。我们的系统目前处于开发初期，但预计在2024年中将具备完备的部署能力。我们将寻找合作伙伴进行商用测试，并不断投入技术研发、拓展市场和提升产品质量。

在销售模式上，我们将采用网页销售、许可证销售、订阅服务、定制化解决方案和授权模式等多种方式。

最后，我们团队将继续努力开发基于人工智能和大数据的技术，为人工智能技术革命贡献一份力量。

“明甄谛听”筑好AI时代的最后一道防线，谢谢大家！