Business Plan for

SocialSafe

Outline

Presentation on YouTube

Business Plan for SocialSafe

Executive Summary

The rise of social media platforms has provided a means for individuals and organizations to communicate and share information globally. However, this has also led to an increase in abusive and harmful content, including hate speech, cyberbullying, and misinformation. To address this issue, automatic tools for detecting and removing abusive content have been developed by social media platforms.

These automatic tools use various techniques, such as machine learning and natural language processing, to identify and remove abusive content. They analyze the content posted on social media platforms and flag potentially harmful content for human review or automatically remove it. This approach helps social media platforms to quickly and efficiently remove abusive content, reducing the harm caused by such content.

Despite their benefits, automatic tools for detecting and removing abusive content also face challenges, such as balancing free speech with harmful content removal and handling cultural and linguistic nuances. Additionally, there is a need for continual improvement of these tools to keep pace with the evolving nature of abusive content.

Overall, automatic tools for detecting and removing abusive content on social media platforms are a promising solution to the growing problem of harmful content. As social media platforms continue to refine and improve these tools, they can help promote a safer and more inclusive online community.

Problem/Opportunity and Solution

Problem/Opportunity:

Social media platforms have become a breeding ground for abusive and harmful content, including hate speech, cyberbullying, and misinformation. Such content can have severe consequences, including harm to individuals and society as a whole. Social media platforms face the challenge of balancing the need to promote free speech with the need to remove abusive content to maintain a safe and inclusive online community. The opportunity lies in developing automatic tools that can detect and remove abusive content efficiently and effectively, reducing the harm caused by such content.

Solution:

Automatic tools for detecting and removing abusive content on social media platforms offer a promising solution to the problem of harmful content. These tools use various techniques, such as machine learning and natural language processing, to analyze the content posted on social media platforms and flag potentially harmful content for human review or automatically remove it. This approach enables social media platforms to quickly and efficiently remove abusive content, reducing the harm caused by such content.

However, developing effective automatic tools for detecting and removing abusive content requires addressing challenges such as balancing free speech with harmful content removal and handling cultural and linguistic nuances. Social media platforms need to continually improve these tools to keep pace with the evolving nature of abusive content.

Company Description

Real.ai is a cutting-edge technology company that specializes in developing automatic tools for detecting and removing abusive content on social media platforms. Our advanced tools use machine learning, natural language processing, and other advanced technologies to analyze the content posted on social media platforms and identify potentially harmful content.

Our mission is to help social media platforms create a safer and more inclusive online community by removing abusive and harmful content quickly and efficiently. We believe that everyone deserves to use social media without fear of harassment or harm, and our tools make this possible.

At **Real.ai**, we are committed to continuous improvement and innovation. We work closely with our clients to understand their specific needs and develop customized solutions that meet their requirements. Our team of experienced professionals has a deep understanding of the social media landscape, as well as expertise in cutting-edge technologies such as machine learning and natural language processing.

We take pride in our ability to provide our clients with reliable, effective, and efficient tools that help them create a positive online experience for their users. At Real.ai, we are dedicated to making the online world a better place, one social media platform at a time.

Industry Analysis

Industry Overview: Automatic Tools for Detecting and Removing Abusive Content on Social Media Platforms

Industry Size, Growth Rate, and Sales Projections:

The industry for automatic tools for detecting and removing abusive content on social media platforms is growing rapidly due to the increasing concerns over harmful content on social media platforms. According to a report by Grand View Research, the market for content moderation solutions, including automatic tools for detecting and removing abusive content, is expected to reach **$11.17 billion** by **2027**, growing at a **CAGR of 10.6%** from **2020** to **2027**.

Industry Structure:

The industry is highly competitive, with a growing number of technology companies entering the market. The industry includes established players such as Jigsaw (owned by Alphabet Inc.) and Two Hat Security, as well as emerging startups such as Real.ai. The industry is characterized by the presence of both large and small companies, offering a range of automatic tools for detecting and removing abusive content on social media platforms.

Nature of Participants:

Participants in the industry include technology companies specializing in developing automatic tools for detecting and removing abusive content, social media platforms, and regulatory bodies. Social media platforms work with technology companies to develop and implement automatic tools for content moderation. Regulatory bodies monitor the industry to ensure that companies comply with regulations governing online content.

Key Success Factors:

Key success factors in the industry include the ability to balance free speech with harmful content removal, the use of advanced technologies such as machine learning and natural language processing, and the ability to handle cultural and linguistic nuances. Additionally, the ability to work closely with clients to understand their specific needs and develop customized solutions is critical for success.

Industry Trends:

The industry is driven by increasing awareness of the negative consequences of abusive content on social media platforms. The industry is also characterized by a focus on innovation, with companies continually developing and refining automatic tools for detecting and removing abusive content. Additionally, there is a growing trend towards collaboration between social media platforms and technology companies to develop more effective tools for content moderation.

Long-Term Prospects:

The long-term prospects for the industry are promising, with social media platforms increasingly prioritizing user safety and working towards creating a positive online experience for their users. The industry is expected to continue growing, driven by the need for social media platforms to address the negative consequences of abusive content. Advancements in machine learning and natural language processing are expected to fuel innovation in the industry, leading to more effective automatic tools for detecting and removing abusive content.

Market Analysis

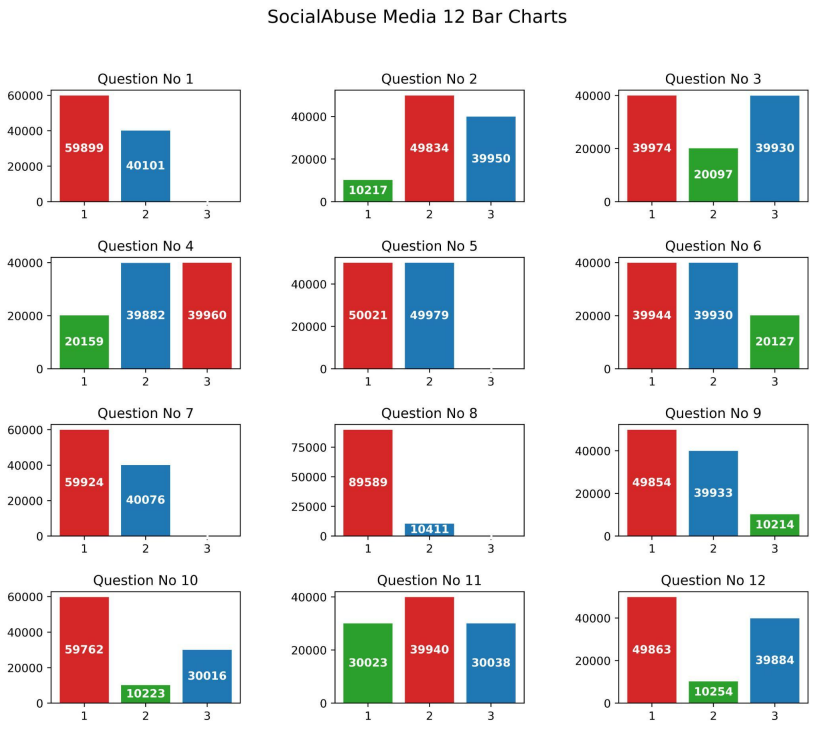
Market Analysis: Automatic Tools for Detecting and Removing Abusive Content on Social Media Platforms

Market Segmentation and Target Market Selection:

The market for automatic tools for detecting and removing abusive content on social media platforms can be segmented based on the type of platform, size of the platform, and geographical location. The target market for these tools includes social media platforms, gaming platforms, online communities, and any other platform that hosts user-generated content. Large social media platforms with a global presence, such as Facebook and Twitter, are the primary target market for these tools.

Buyer Behavior:

Buyer behavior in the industry is driven by concerns over the negative consequences of abusive content on their platform, such as user safety, reputation damage, and regulatory compliance. Social media platforms prioritize user safety and work towards creating a positive online experience for their users. Therefore, they are highly motivated to invest in automatic tools for detecting and removing abusive content on their platforms.



Google Form
https://docs.google.com/forms/d/e/1FAIpQLSfwGSJ7bTNB98kCk8veCqR_ECyZ29rkdDqB7t9IDrZP8ZUV0w/viewform?usp=sf_link

- Approximately 60 % of the people are victims of social abuse so our service of ***Abuse Detection and Removal Bot*** will be purchased by a lot of people.
- Which is Abuse is **Rated** the most

Spreading False Rumors / News	10 %
Offensive Comments	50 %
Unappropriated Media *(Audios, Videos)*	40 %

Here the Comments and Media are most disturbing for Social media Users. So they need something that they do not hate to get to them which is the primary reason for developing this bot.

- From the 3rd Question we can conclude that the ones who are getting hate are more frequently getting it so it is not just a need it is a necessity for them.
- Most of the time the reply from the help desk is not that helpful. Especially when you can't even leave that group of chat in which you are hated. Our Bot will rephrase it so that you won't feel bad about it.
- The answer shows that most of the time this exchange of words could lead to real life fighting ***normal among teenagers***.
- Not just individuals but their colleagues also witnessed it but could not do anything as they felt weak. This weakness is because of continued hate.
- Most people have limited social media usage because of this ***Rumors***, ***Offensive Comments*** and ***Unappropriated Media*** (*Audios*, *Videos*).

  If the Social Media company notices it this will be a great loss to lose customers at such a high rate.

- The customers are quite comfortable if a bot filters hate for them and only delivers them with useful information.
- If you want to send such a message of hate to others but you just want them to know that you also feel that you are being angry at their act.

  The **Bot** will do it but in a manner it would express yourself and your message very carefully.

Analyzing Expert Questionnaire:

- The quality of the ***training data*** is crucial to the effectiveness of the model.

- Another approach is to involve human experts like ***linguists***.
- This can be done through continuous monitoring and analysis of the bot's performance with the help of ***Reinforcement Learning***.
- One Major Concern: Diversity and ***Complexity*** of language use on social media platforms.
- Bot's performance depends on ***Data change on media*** and ***Linguists marking their decision***.
- The Real World performance will be evaluated from user experience through ***feedback***.
- Our form will contain fields for their expectation and what bot has generated and then it will be delivered to linguists for processing.
- Limiting access to sensitive data and regularly auditing data handling practices with no label of user info for performance improvement.

Competitor Analysis:

Although the competitors listed are currently offering solutions in the content moderation market, their primary focus is on serving businesses and organizations that are looking to moderate content on their platforms. As such, their target market does not directly overlap with Real.ai's target market, which is focused on providing content moderation solutions for individual social media users. While there may be some overlap in terms of the technology and capabilities required to develop effective content moderation solutions, Real.ai's focus on providing a tailored, modular platform for individual social media users sets it apart from its competitors.

1. **Perspective API by Google:** Perspective API is an API that uses machine learning to identify toxic language and provide feedback to users. It is currently used by various media organizations and social media platforms.
2. **Two Hat's Community Sift:** Community Sift is a content moderation platform that uses artificial intelligence and machine learning to detect and remove harmful content. It is currently used by several gaming and social media companies.
3. **Sift by Sift Science:** Sift is a fraud prevention and content moderation platform that uses machine learning to detect and prevent harmful content. It is currently used by several e-commerce and financial services companies.
4. **Spectrum Labs:** Spectrum Labs is a content moderation platform that uses machine learning and natural language processing to detect and remove harmful content. It is currently used by several social media and gaming companies.

Competitor	Perspective API	Two Hat's Community Sift	Sift by Sift Science	Spectrum Labs	Real.ai
price	Free for limited use, pay-per-use for higher volumes	Custom pricing based on usage	Packages like normal and premium	Custom pricing based on usage	Free for limited use, pay-per-use for higher volumes
Availability	Available to select partners and in certain languages only	Available but not in pakistan	Available but not in pakistan	Available but not in pakistan	Available
Speed of service	Fast	Fast	Fast	Fast	Fast
Reliability	High	High	High	High	High
Familiarity	Widely known	Just famous in West	Well-known in the fraud prevention industry	Relatively new in the market	New entrant in the market
Customer Trust	High	High	High	High	High
Strengths	Strong brand recognition as a Google-backed platform	Comprehensive moderation platform with robust AI and machine learning capabilities	Scalable platform with strong fraud prevention capabilities	Comprehensive content moderation capabilities with strong AI and natural language processing capabilities	Customizable, modular platform with advanced natural language processing capabilities
Weaknesses	Limited customization options and only detects toxic language	Primary focus on gaming industry	Limited focus on content moderation	Smaller customer base and relatively new in the market	Limited brand recognition as a new entrant in the market

Estimates of Annual Sales and Market Share:

According to a report by Grand View Research, the market for content moderation solutions, including automatic tools for detecting and removing abusive content, was valued at $5.45 billion in 2019. The same report projects the market to reach $11.17 billion by 2027, growing at a CAGR of 10.6% from 2020 to 2027. As for market share, the industry is highly fragmented, with no single company holding a significant share of the market. However, established players such as Jigsaw and Two Hat Security are expected to maintain their position due to their strong brand reputation and established customer base.

Marketing Plan

Key partners are individuals or organizations that a business works with to achieveThe overall marketing strategy for automatic tools for detecting and removing abusive content on social media platforms should focus on building brand awareness, establishing credibility, and targeting key decision-makers at social media platforms and online communities. This can be achieved through a combination of content marketing, search engine optimization (SEO), and social media advertising.

Product:

The product offering should include automatic tools for detecting and removing a wide range of abusive content, such as hate speech, cyberbullying, and fake news. The tools should be customizable to meet the unique needs of each platform and integrate with existing content moderation systems.

Price:

Pricing for automatic tools for detecting and removing abusive content will depend on several factors, including the size of the platform, the number of users, and the level of customization required. A subscription-based pricing model may be appropriate, with pricing tiers based on the size of the platform.

Promotions:

Promotions should be focused on highlighting the benefits of automatic tools for detecting and removing abusive content, such as increased user safety, improved platform reputation, and compliance with regulatory requirements. Promotional tactics may include targeted advertising on social media platforms, attending industry conferences, and partnering with industry influencers.

Distribution:

Distribution of automatic tools for detecting and removing abusive content will primarily be through direct sales to social media platforms and online communities. However, partnerships with third-party content moderation providers may also be considered.

Sales Process:

The sales process for automatic tools for detecting and removing abusive content will typically involve a needs assessment, customization of the tools to meet the unique needs of the platform, and ongoing customer support.

Sales Tactics:

Sales tactics for automatic tools for detecting and removing abusive content should focus on establishing credibility and building trust with key decision-makers at social media platforms and online communities. This can be achieved through the use of case studies, testimonials, and demonstrations of the product in action. Additionally, offering free trials or demos may be an effective tactic for generating interest and driving sales.

Design and Development Plan

Development Status and Tasks:

The development status of automatic tools for detecting and removing abusive content on social media platforms may vary depending on the company offering the product. However, the main tasks involved in the development of such tools may include identifying and analyzing different types of abusive content, designing and training machine learning models, integrating the tools with existing content moderation systems, and conducting regular updates and maintenance.

Challenges and Risks:

One of the main challenges and risks associated with developing automatic tools for detecting and removing abusive content is ensuring the accuracy and fairness of the tools. Machine learning models may be biased or inaccurate, which can lead to false positives or negatives, resulting in the removal of legitimate content or the failure to remove abusive content. Additionally, ensuring compliance with privacy laws and regulations while monitoring user-generated content is another challenge and risk that companies in this space must navigate.

Projected Development Costs:

The projected development costs of automatic tools for detecting and removing abusive content on social media platforms will depend on the complexity of the product, the size of the platform, and the level of customization required. However, development costs may include salaries for software engineers and machine learning experts, research and development costs, and the cost of acquiring or licensing data sets.

Proprietary Issues:

Companies offering automatic tools for detecting and removing abusive content on social media platforms must ensure they have proper intellectual property protections in place. This may include patents, trademarks, copyrights, licenses, and brand names. Additionally, companies must be careful not to infringe on the intellectual property of other companies or individuals, as this can lead to costly legal battles and damage to the company's reputation.

Operational Plan

General Approach to Operations:

The operations plan for automatic tools for detecting and removing abusive content on social media platforms will involve ongoing monitoring, maintenance, and updates to ensure the accuracy and effectiveness of the tools. This includes regularly updating machine learning models and algorithms, monitoring performance and accuracy, and addressing any issues or bugs that arise.

Business Location:

The business location for developing automatic tools for detecting and removing abusive content on social media platforms may vary depending on the company's size and needs. However, a location with access to technology and machine learning expertise is essential. Additionally, proximity to social media platforms and potential clients may be beneficial.

Facilities and Equipment:

The facilities and equipment needed for developing automatic tools for detecting and removing abusive content on social media platforms may include:

- **Office Space:** A space for the development team to work together, collaborate and hold meetings.
- **Computer and Servers:** A reliable computer system and servers are needed to store data and run the machine learning models.
- **Software and Tools:** Software and tools for machine learning, data analysis, and programming are necessary.
- **Communication Tools:** Collaboration tools such as instant messaging, video conferencing, and email are necessary for effective communication.
- **Security and Privacy Measures:** Since the tools will be dealing with sensitive data, robust security and privacy measures are essential to protect data and ensure compliance with privacy laws and regulations.

Projected Costs:

The projected costs for the operations plan of automatic tools for detecting and removing abusive content on social media platforms will depend on the size of the team, the complexity of the product, and the length of the development cycle.

However, costs may include salaries for developers and support staff, office rent, server maintenance, and security measures.

Risk Mitigation:

To mitigate risks associated with operations, companies should establish clear processes for maintaining and updating the tools, and ensure that the tools are always up-to-date and effective. Additionally, companies should have contingency plans in place for unexpected issues that may arise during operations, such as server downtime or security breaches.

Management Team and Management Structure

Management Team:

Real.AI's management team consists of experienced professionals with a range of skills in technology development, product management, and marketing. With the CEO/CTO Muhammad Hassan Mukhtar being 5 years of AI Field Experience.

Board of Directors:

Real.AI's board of directors includes experienced professionals in technology, finance, and law. The board of directors includes:

- A technology executive with over 20 years of experience in the industry. He has held executive positions at several leading technology companies.
- A seasoned venture capitalist and has served on the board of several successful startups.
- A lawyer with extensive experience in corporate law and has worked with several leading technology companies.

Board of Advisors:

Real.AI's board of advisors includes experts in technology development, product management, and marketing. The board of advisors includes:

- A leading expert in machine learning and has published several papers in top-tier academic journals.
- A marketing executive with extensive experience in product marketing and has worked with several leading technology companies.

Company Structure:

Real.AI's company structure is organized around a product-focused approach, with teams dedicated to developing and maintaining the company's automatic tools for detecting and removing abusive content on social media platforms. The company structure includes:

- **Product Development Team:** This team is responsible for developing and maintaining the company's machine learning models and algorithms.
- **Operations Team:** This team is responsible for monitoring the performance of the tools and ensuring that they are always up-to-date and effective.
- **Sales and Marketing Team:** This team is responsible for promoting the company's tools to potential clients and generating new business opportunities.
- **Administrative Team:** This team is responsible for handling day-to-day administrative tasks, including accounting, human resources, and legal.

Financial Plan

Cost Structure:

When starting and operating a business, there are various costs that need to be considered, including human resource costs, technical resource costs, software development costs, legal and regulatory costs, and marketing and promotion costs.

- **Human Resource Costs** refer to the costs associated with hiring and maintaining a workforce. These costs include salaries, benefits, and expenses such as furniture, lighting, and airflow. The cost of salaries is typically the largest expense, as it represents the primary cost associated with employing staff. The cost of furniture, lighting, and airflow is also significant, as these expenses impact the comfort and productivity of the workforce.
- **Technical Resource Costs** refer to the expenses associated with acquiring and maintaining the hardware and software needed to operate the business. This includes purchasing desktops for developers and servers to host the service. These costs can be significant, particularly in the early stages of the business when infrastructure needs to be established.
- **Software Development Costs** refer to the expenses associated with building and maintaining the software needed to deliver the service. This includes the cost of software development tools, licensing fees, and the cost of hiring developers and engineers.
- **Legal and Regulatory Costs** refer to the expenses associated with complying with local laws and regulations. This includes the cost of legal advice, compliance audits, and regulatory fees.

Revenue Streams:

When it comes to generating revenue for a language modeling service, there are various options that can be explored. One option is to charge a subscription fee to users who wish to access the service. This can be an effective way to generate revenue, particularly if the service is highly specialized and caters to a specific market.Following will be some ways to generate revenue from our service:

- A subscription fee is a type of revenue model that involves charging customers a recurring fee to access a service or product. This fee can be charged on a weekly, monthly, or yearly basis, and can be an effective way to generate consistent revenue over a period of time.

- Adopt a commission-based model where the service takes a percentage of the revenue generated by the platform or advertisers who use the service. This model can be particularly effective for language modeling services that are integrated into other platforms, such as social media platforms or e-commerce websites.
- The data generated by the language modeling service can also be a valuable source of revenue. By selling unlabeled data for business analysis, the service can generate additional revenue while also helping businesses gain insights into their customers' language usage patterns.
- Another way to generate revenue is to arrange workshops in research institutes and other academic settings. These workshops can be used to introduce researchers and linguists to the language modeling service and its capabilities, which can lead to collaborations and partnerships that generate revenue for the service.

Overall Scheduling

Following is an overview of how this project will be developed:

Phase	Time	Description
Research and Development:	6 months	Developing algorithms and models for identifying abusive content using natural language processing, machine learning, and other technologies.
Data Collection and Analysis	4 months	Collecting and analyzing data on abusive content, including user reports and feedback, to train and refine the algorithms used by the automatic detection tools.
Testing and Validation:	5 months	Testing and validating the effectiveness of the automatic detection tools using real-world scenarios and data sets.
Integration with Social Media Platforms	3 months	Integrating the automatic detection tools with social media platforms and APIs to enable automatic detection and removal of abusive content.
Monitoring and Evaluation	ongoing	Monitoring the performance of the automatic detection tools, evaluating their effectiveness in

		identifying and removing abusive content, and making necessary adjustments to improve performance.
Collaboration and Partnerships:	ongoing	Collaborating with social media platforms, technology companies, nonprofit organizations, government agencies, academic researchers, and user advocacy groups to develop and implement effective automatic tools for detecting and removing abusive content.
Education and Awareness:	ongoing	Educating users about the importance of online safety and promoting awareness of the automatic detection tools and how they can be used to combat abusive content on social media platforms.

** All these phases will form one cycle and there will be 3 cycles for three types of media text, audio and photos/videos.*