May 01, 2023

sonata

WHITEPAPER

Version 0.92.8

Table of Contents

Section 1 Introducing the problem and the solution

Section 2 Defining The Practical Ideal – Introducing 3rd Generation Social Media

Table of Contents

Section 3 Philosophy and Approach of Sonata

Table of Contents

Section 3 Philosophy and Approach of Sonata (Cont.)

Introduction

Microblogging has rapidly grown into a hugely impactful medium, reaching billions of people worldwide each month. From its roots as a mode of public expression and communication between people of any background, it has also developed into a systematic utility for self-serve broadcasting of news, opinion and knowledge from the world's most influential, powerful and notable people and organisations.

The meteoric rise of such an impactful new system has created many issues, criticisms and problems, many of which have been left unaddressed or have subsequently increased in severity.

This whitepaper outlines these issues and presents solutions, leading to the introduction of a new microblogging platform, Sonata, due for early release shortly after this paper. Sonata has been designed and built from the ground up with these issues, concerns and failings addressed on a structural and ideological level into a system that as its very core, seeks to make consistently positive impact on society through carefully constructed policies, reliable management, resistance to outside influence and the constant pursuit of constructive innovation.

It also introduces the generic concept and definition of a "third generation" social media platform, which outlines the rules and standards that any platform and operator must follow to be considered part of the latest developments in responsibility, conduct and contribution to society.

Glossary and Definitions

Microblogging	Posting text based content to an online platform which maintains a small character limit per post, with or without accompanying media.
Person	A human. Defined here so as not to conflate with 'user'
User	An operator of an account on a platform that is assumed, yet not been fully confirmed, to be a person.
Attention Economy	A model typical of social media which prioritises users' continued attention, usually to maximise revenue through advertisements.
Illegal Content	Content that is not legal to promote or provide within the country of Iceland[1]. Special consideration is also taken for content that is not legal in the United States of America and the European Union; however, these may be rescinded if future laws pushing back against the principles of free speech are brought into force.
User Responsibility	A platform carries a duty to all people using its service to provide a secure, stable and reliable service that befits the trust and commitment they have made to the service. It must recognise that it has been entrusted with personal data that users expect to be kept safe.

A platform is also entrusted with the ability to push information onto considerable numbers of people, often openly used as the primary source of monetisation, which, if used, should be recognised as a privilege.

A person expressing their opinions on the platform will be seen to have done so freely and of their own accord; a platform must not abuse this trust by amending a message covertly. It must also be careful not to promote content in such a way that users have their expressions taken out of context. Removing messages as part of content moderation can also cause this effect, so a display should be left to show that missing related content was deleted.

Social Responsibility

A platform facilitating, managing and promoting the interactions of a large number of people has a huge responsibility to act diligently in the interests of each person and society as a whole.

While everyone should assume their right to freedom of expression, they should also be protected from unwanted exposure to material created by those who wish to influence or manipulate them, especially where such material would be considered harmful, dangerous or inappropriate.

To achieve and uphold social responsibility, the following must be understood and accounted for by social media platforms:

- Each person must be recognised as a human being, with rights bestowed upon them by the UN Charter of human rights.
- Each person should be able to expect a fair experience from their interaction with the platform.
- Accessibility must allow all of age people to use the platform, within reason.
- Minors should benefit from extra protections kept in place until they turn 18.
- People are vulnerable to manipulation, as is built into our psychology, and this should not be exploited for the benefit of the platform.
- Some people will try to abuse the platform for their gain and will try to hide this abuse to continue.
- Potential for mental harm can arise from the use of social media platforms. Signs of this (such as overuse) should be addressed, and assistance should be offered.
- A person's opinion is just as valid as anyone else's, regardless of the number of supporters each has.
- Echo chambers insulate and divide, preventing natural discourse and discussion. Avoiding the development of echo chambers must be strived for.

Microblogging Related Problems

Current platforms have many fundamental issues, the most pressing of which are identified below, along with practical solutions

Classification

The problems with microblogging and currently available platforms can be divided into several categories: societal, individual and management.

Individual issues relate to topics including:

- The technological understanding and accessibility of the platform.
- The risk of being influenced into harmful or dangerous decisions or opinions.
- Impacts on mental health.
- The isolated, targeted or disproportionate suppression of expression.

Societal issues relate to topics including:

- The impact on the natural progression of public discourse.
- Dissemination and flow of information, factual or otherwise.
- Disproportional representation of race, ethnicity, background or culture.
- Exposure of dangerous material to those most vulnerable.

Management issues relate to topics including:

- Platform control and top-level decision making.
- The burden of expenses to enable maintenance of social responsibility.
- Pressure from financiers to make changes to policies.

Issues & Idealised Solutions

Individual Issues

A platform holds unregulated power over the ability of people to express themselves. They have the ability to censor, suppress, misrepresent or de-platform without regulatory oversight.

Content policies must be comprehensive and exhaustive. Changes to policies must be broadcast publicly. Moderation contrary to content policies must be expressly disallowed and called out by independent parties. Data must be publicly released as much as possible for independent review.

The attention economy can lead to the disproportionate dissemination of divisive content, as conflict and controversy can have a higher likelihood of holding users' attention for a greater length of time, which can lead to measurable harm to the individual.[2]

The attention economy should be rejected, and algorithmic content recommendation systems should be avoided. Potentially harmful situations (such as 16-year-olds spending 6 hours on the platform on a weekday evening) should be identified and reacted to responsibly (such as reminding them of the time spent or creating systems designed to help them take a break).

Societal Issues

Algorithmic content recommendation helps to create and encourage echo chambers in which opinions are often radicalised and dissenting views are obscured, allowing many people to develop excessively narrow-minded viewpoints.

Algorithmic content recommendations should be removed. If required, recommendation engines should be developed that are specifically designed to avoid the creation of harmful echo chambers.

Societal Issues (Cont.)

Misinformation is often easier to spread than information (due to the attention-grabbing effect of fear and readily-given answers and the added difficulty of fact checking). Labelling content in a 'bottom-up' approach[3], as is becoming commonplace, takes excessive time, effort, and money. It is also hard to deal with without creating censorship issues.

A recent study on Covid-19 misinformation has shown[4] that the majority of the content can be originally created and spread by a tiny group. If prevalent elsewhere, working to identify and demote material (as recommended by Freedom House) created by these groups in a top-down approach will significantly reduce its spread.

Freedom of expression is often at risk as platforms seek to maximize revenue, influencing decisions and policies. Platforms may also overdo content moderation, amounting to platform-mandated censorship to ensure they isolate themselves from legal liability

Revenue should be mandated through platform policy as a lower priority to freedom of expression. Platforms should be legally protected while allowing freedom of expression within legally permitted limits (For example, distributing CSAM[5] content breaches the law of nearly every country, so it would not be lawfully permissible freedom of expression). Suitable tools and moderators should be employed to employ content policies more carefully. An appeal system should always allow a person to easily challenge a decision.

The suppression of journalists is a dangerous affront to freedom of information and democratic procedure.

Journalists[6] should be able to apply for a specialised account tag, protecting them from automated moderation, user report spam causing automatic limitations, and targeted suppression from biased admins or moderators. They should also be given special protections from censorship, such as pressure from state actors. It is also essential that they can choose to be anonymous, such that their personal safety is not at risk due to their use of the platform.

Societal Issues (Cont.)

Minorities suffer de-facto suppression, beyond that of simple proportion, due to the naturally taken actions of the majority. The inability to gain sufficient recognition is one part of this issue. Dismissive or hostile responses due to cultural differences, leading to the shaping of general opinion through 'social proof' and influence, is another.

Platform accessibility is paramount for all user needs. Native platform translation should be provided for all needs when financially possible. Content translation should be provided where possible. Internal investigations should be carried out on minority groups gaining little exposure. External studies on the same topic should be aided and permitted. Negative systematic responses (such as dislikes) should be made less accessible from the general public's view.

Users under 18 are at an exceptionally high risk of harm from harmful or manipulative content. They are at higher risk from other platform users for behaviours such as grooming. They are also at elevated risk from using a platform due to overuse.

Users under 18 should be required to use an account type designed for their age. However, this should be done without impact to adult users, who should not be required to expose their personal sensitive documents, such as through uploading their ID. The child account should limit the ability to access harmful content and interactions between children and adults. Parents or guardians should be given partial access and control over their child's actions and activity, with some actions staying private, such as interaction with LGBTQIA+ communities.

Management Issues

Sources of revenue, such as advertisers or licenced providers, may threaten to withdraw or otherwise withhold their business if content policies are not changed to suit them, which threatens freedom of expression.

Alternative concurrent sources of revenue should be sought that do not have this issue. Advertisers should be pressured back to reduce these demands.

Management Issues (Cont.)

Global platforms for hundreds of millions of users require large operations with large financial overheads that cannot quickly be reduced. This may lead to abrupt policy changes if the necessary level of revenue is threatened or diminished.

Multiple revenue sources should be sought and maintained. Reserves of money should be kept to allow the company to continue if revenue unexpectedly drops. Prudent staffing should be maintained to reduce wasted expenditure.

Market Overview

Summary

The microblogging market share is exceptionally top-heavy[7] , with almost all users on the most popular platform, Twitter. Within the past year, some new platforms have appeared in response to many users quitting Twitter due to recent changes in management and policy. Other platforms have emerged either as general competitors or for subsets of users that disagree with Twitter's content policies.

Due to the turmoil of Twitter's recent changes, many users are trying other platforms, only to leave them shortly after, making it difficult to obtain accurate figures for monthly average users.

Twitter

Release: 2006
Average Monthly Users: 556,000,000[7]
Effective Audience: Global
Primary revenue sources: Advertising, paid subscription
Base software: Proprietary
Ownership Structure: Privately Owned
Publicly Accessible Without Login or App Download: Yes
Assessment of Key Platform Issues:

- New ownership causing anger and uncertainty.
- New features and policies have upset users, poor reception has led to many turnarounds.
- Content policies are becoming more restrictive.

Mastodon

2022 saw the rise in popularity of federated microblogging system Mastodon. Due to its nature as a self-hosted platform, in addition to considering the software as a whole for monthly active users, each implementation should be considered separately.

Mastodon (Cont.)

Release: 2016
Average Monthly Users: 1,200,000[8]
Highest Monthly Users Per Instance[9]: 170,000[10]
Effective Audience: Global
Primary revenue sources: Developers supported via Patreon; individual instances may be self-funded or supported by advertising.
Base software: Open Source
Ownership Structure: Privately Owned
Publicly Accessible Without Login or App Download: Some instances yes
Assessment of Platform Issues:

- More complex than other platforms to sign up for and use.
- Instances cause a natural split in the user base, as content does not carry between them. Natural echo chambers are formed.
- Content policies are different for each instance and may confuse. The burden of moderation on a hobbyist owner may cause restrictions to user numbers and heavy-handed arbitrary moderation.

Hive Social

Release: 2019
Average Monthly Users: Estimated 200,000 - 800,000[11]
Effective Audience: Global
Primary revenue sources: Advertising
Base software: Proprietary
Ownership Structure: Privately Owned
Publicly Accessible Without Login or App Download: No
Assessment of Platform Issues:

- Non-unique handles, making impersonation much more likely.
- Privacy concerns, openly shares personal data with third parties.
- 'Hive' is a very commonly used name and is causing confusion, particularly with a decentralised blockchain protocol of the same name.

Threads

Release: 2023
Average Monthly Users: 10m+ (dropping since launch)
Effective Audience: Global
Primary revenue sources: Advertising.
Base software: Proprietary / ActivityPub (coming soon)
Ownership Structure: Privately Owned
Publicly Accessible Without Login or App Download: No
Assessment of Platform Issues:
• Not available in the EU due to privacy concerns.
• Very restrictive content policy.
• Sigificant data concerns.

Bluesky

Release: 2023 (Invite only as of August 2023)
Average Monthly Users: 100,000 (estimated)
Effective Audience: Global
Primary revenue sources: None
Base software: Open Source
Ownership Structure: Privately Owned by a small group, partially investor funded
Publicly Accessible Without Login or App Download: No
Assessment of Platform Issues:
• Untested federated protocol.
• Poor funding outlook as they have committed to zero ads.
• Untested at scale crowd moderation system.

Tribel

Release: 2017
Average Monthly Users: Unknown, assumed 250,00 - 1,000,000
Effective Audience: Global
Primary revenue sources: Self funded, advertisement planned
Base software: Proprietary
Ownership Structure: Privately Owned
Publicly Accessible Without Login or App Download: Yes
Assessment of Platform Issues:

- Aggressively left-leaning ideology, with other views blocked by the platform.
- It has been criticised for requiring users to agree to exploitative terms and conditions.

Introduction

Continuously developing since the early 2000s, social media has rapidly expanded into the society-defining behemoth it is today. During this time, two distinct generations of platforms have emerged, the early exploration generation and the current attention-based generation that has reigned since the early 2010's.

This whitepaper presents the concept of a new "third-generation" social media platform, which, along with each other distinct generation.

Generation 1: Discovery

Past examples (as they were in 2008): SixDegrees.com, Myspace, Bebo

Definition: A platform offering social interaction as a primary function but not engaging in research and techniques to maximise user engagement and revenue.

Generation 2: Attention

Current examples: Facebook, Twitter, Instagram, Tiktok

Definition: A platform that bases its model around collecting and using user data for financial gain and utilises algorithmic content recommendation techniques to maximise and prioritise the length of time a user spends on the platform.

Generation 2.5: Opposition

Current examples: Mastodon, Hive, Tribel

Definition: A platform which rejects the exploitation of user data, the attention economy or any other facet of existing platforms and is definable by this opposition. For example, Mastodon rejects the centralised nature of previous-generation platforms and the difficulty of switching between platforms.

Generation 2.5 platforms may adapt over time to embrace either the principles of either 2nd or 3rd generation platforms, changing their classification.

Generation 3: Social Responsibility

Current examples: None

Definition: A platform that prioritises the betterment of society, fully commits to the UN Charter of human rights, protects its users from censorship and harmful effects of social media and follows the 33 definitions of a third-generation social media platform given below.

Considerations

The full definition of 3rd generation social media must be all-encompassing for all types of platforms, not just microblogging. It should also be practical for a controlling party to achieve and maintain.

The definition should be subject to broader scrutiny following the release of this whitepaper and should be amended accordingly before being published as a first version. Following this, it should be publicly scrutinised after additional exposure, leading to a finalised version.

Definition

1. Prioritise users' free expression and access to information, particularly for journalism; discussion of human rights; educational materials; and political, social, cultural, religious, and artistic expression.
2. Prioritise and set as a continuing goal the improvement of society through top-level platform decisions, policies, actions and innovation.
3. In guidelines and terms of service, clearly and thoroughly explain what speech is not permissible, what aims restrictions serve, and how content is assessed for violations. Ensure that terms of service and mechanisms for reporting harmful content and appealing content decisions are translated into all languages where the company's products are used.
4. Reject the attention economy and personalised algorithmic content recommendations.
5. Address the United Nations Bill of Human Rights and detail how the service will conform with it in all relevant articles.
6. Address corporate social responsibility and maintain a policy in respect of this.
7. Require only the necessary licenses for user-supplied data and content which permit the service to perform its primary purpose of storing, aggregating and distributing such material in a way that remains justifiably financially viable. Any further licences must be optional, opt-in and separately agreed upon.
8. Provide complete transparency regarding published policies. Detail all policy items in extreme detail. Reasoning should include the source of all rules, definitions of all potentially ambiguous terms and case studies or examples for any complex policy item.
9. Publish detailed transparency reports on content removals, both for those initiated by governments and for those undertaken on behalf of individuals or companies. If applicable, transparency reports should also address how machine learning is used to train automated systems that classify, recommend, and prioritize content for human review.

Definition

10 Publish regularly updated statistics on data relating to users of the platform by month, in total and total by country:
- Active users
- Total users
- Post creation
- Total views and engagements
- Number of people whose private data is being held

11 Publish as much data as possible without breaching privacy or security responsibilities and justify each restriction.

12 Permit a yearly independent investigation and subsequent report into the company's internal affairs, including:
- Legal matters (active case information may be redacted if advised)
- Finances
- Data protection & privacy
- Platform security

13 Open source all code which has involvement in the following:
- Content aggregation (including feeds)
- Content and user search
- Content and user discovery features (such as trending lists)
- Processing of data relating to users
- Content recommendation

14 Establish and publicise a full disclosure on platform ownership, data storage and data processing locations and responsibility for all platform aspects.

15 Openly make available to any self-identifying person over the age of 18 who makes a formal request:
- All publicly available content posted by any person.
- All publicly available data relating to any person.

16 Ensure an exceptionally high standard of data protection and privacy by:
- Developing security measures to prevent data leaks.
- Resisting requests for private information to be revealed by courts, state agencies and governments unless the user's actions on the platform constitute a major crime in the platform's jurisdiction.
- Removing or anonymising personal data is that no longer required.
- Requiring strong security measures for any administrator able to access personal data.
- Appropriately managing administrator access rights to limit exposure of data if a leak occurs.

17 Allow for and support fully anonymous users.

18 Actively and proactively prevent the exploitation of the platform to manipulate political elections.

19 Review and verify the truthfulness of all political advertisements.

20 Actively provide preventive guidance for those searching for topics dangerous to mental health or likely to cause direct physical harm to the user searching.

21 Actively provide functionality to aid those with suspected social media related or induced issues, such as social media addiction or social media induced depression.

22 Reject the use and support of tools that facilitate and aid the propagation of unrealistic personal appearances.

23 Provide optional functionality to reduce or restrict access to types of content as the user requests.

24 Provide an efficient and timely avenue of appeal for users who believe their rights were unduly restricted, including through censorship, banning, assignment of labels, or demonetization of posts.

25	Facilitate and permit the existence of third-party clients in a financially viable way. If charging a fee, break down the costs in detail and justify.
26	Promote healthy competition by interworking services with competitors.
27	Ensure continued value of investment in the event of platform decline.
28	Expand the capacity, geographic, and linguistic diversity of content moderation teams, and ensure they are sensitive to nuances in a language that is spoken across multiple countries or regions. Conduct human rights due diligence assessments to ensure that implementation of moderation does not lead to unintended consequences, such as disproportionately affecting marginalized communities.
29	Ensure advertisements cannot be tailored specifically to single users, especially through the use of AI.
30	Make freely available to anyone self identifying as a researcher any content which was removed as a result of moderation action, along with the reasons for removal, provided you are legally able to do so. If possible, redact non legal text based content sufficiently that it can be made available.
31	Provide a warning of no less than 7 days to relevant parties about any new factual labels which are to be applied publicly to an account representing an institution or business. Facilitate a system of appeal, allowing the entity to provide evidence that the label is inaccurate and should be amended, or withdrawn. When publishing a label, publicise relevant data concerning the decision including data from any appeals, if the entity grants permission. Ensure that all labels applied are accurate to the best of the platform's knowledge. Permit appeals from the entity at any later time for further amendment.
32	Allow any user to close their account and delete all personal data, without any consequence to any unrelated services.
33	Resist governmental pressure from any country on any issues that degrade these policies.

Definitions Comparisons Chart

With these definitions established, it becomes possible to gain insight into the current position of existing platforms, with respect to their position on social responsibility. All scores were derived from research into each platform and may be subject to updates following feedback.

All of the following scores and notes are the opinion of Sonata Social ehf, obtained through research into each social platform. If anyone wishes to provide an evidence backed suggestion to this information, we will make appropriate updates in later versions of this whitepaper.'

Scoring system works like this:

- **-1:** Platform explicitly goes against the requirement
- **0:** Platform does not support or mention the requirement, or appears to act in contrast to stated support.
- **1:** Platform appears to support the requirement, but provides no firm evidence of this, or appears unable to consistently commit to meeting it.
- **2:** Platform partially supports the requirement and appears to be improving
- **3:** Platform fully and explicitly supports the requirement

Microblogging Platforms

3rd Generation Definitions	Twitter	Mastodon*	Hive	Tribel	Threads	Bluesky
1 Prioritise users' free expression and access to information ...	1 *	1 *	1 *	0 *	2 *	2
2 Prioritise and set as a continuing goal the improvement of society ...	1	3	1 *	0	0	3 *
3 Clearly and thoroughly explain what speech is not permissible ...	1 *	1 *	0 *	1 *	2 *	1 *
4 Reject the attention economy and personalised algorithmic ...	-1 *	2	3	1 *	-1	-1 *
5 Address the United Nations Bill of Human Rights and detail how ...	0	0	0	0	0	0
6 Address corporate social responsibility ...	0	0	0	0	3	0
7 Require only the necessary licenses for user-supplied data ...	0 *	1 *	-1 *	2	-1 *	1 *

Definitions Comparisons Chart (Cont.)

3rd Generation Definitions	Twitter	Mastodon*	Hive	Tribel	Threads	Bluesky
8 Provide complete transparency regarding published policies ...	0 *	0 *	0	1	1	1
9 Publish detailed transparency reports on content removals ...	1 *	0	0	0	2 *	0
10 Publish regularly updated statistics on data relating to users ...	0 *	2	1 *	1 *	1 *	1 *
11 Publish as much data as possible without breaching privacy ...	1 *	2	0	0	0	0
12 Permit a yearly independent investigation and subsequent ...	0	2	0	0	1	0
13 Open source all code which has involvement in the ...	-1 *	3	0	0	0	2 *
14 Establish and publicise a full disclosure on platform ownership...	0	1 *	1 *	1 *	3	1
15 Openly make available to any self-identifying person over the age of...	0 *	1	0 *	0	0	0
16 Ensure an exceptionally high standard of data protection ...	1	2	2	2	2	2
17 Allow for and support fully anonymous users.	2	2	1 *	2	2	2
18 Actively and proactively prevent the exploitation of the platform to...	3	1 *	0	0	0 *	2
19 Review and verify the truthfulness of all political advertisements.	3 *	3	0	-1 *	1	3
20 Actively provide preventive guidance for those searching for ...	0	0	0	0	2	0
21 Actively provide functionality to aid those with suspected social ...	0	0	0	0	1 *	0
22 Reject the use and support of tools that facilitate and aid the ...	0 *	2 *	2 *	2	0	2

Definitions Comparisons Chart (Cont.)

3rd Generation Definitions	Twitter	Mastodon*	Hive	Tribel	Threads	Bluesky
23 Provide optional functionality to reduce or restrict access to types...	3	3	2 *	0	1 *	1 *
24 Provide an efficient and timely avenue of appeal for users who ...	1 *	0	0 *	0 *	0 *	2
25 Facilitate and permit the existence of third-party clients in a ...	-1 *	3	0	0 *	0	3
26 Promote healthy competition by interworking services with ...	0	3	0	0	1 *	3
27 Ensure continued value of investment in the event of ...	0	3	0	0	0 *	3
28 Expand the capacity, geographic, and linguistic diversity of content ...	2 *	0	0	1	2	1
29 Ensure advertisements cannot be tailored specifically to single ...	0	3	2	1 *	1	3
30 Make freely available to anyone self identifying as a researcher ...	0	0	0	0	0	0
31 Provide a warning of no less than 7 days to relevant parties about ...	0	2 *	2	2 *	2 *	2
32 Allow any user to close their account and delete all personal ...	3	1 *	3	3 *	-1 *	2
33 Resist governmental pressure from any country on any issues that ...	2	2 *	0	0	0	0
Maximum Available Points 99	21	49	20	19	27	42

*Mastodon is assessed here as an average of the largest instances

Definitions Comparisons Chart (Cont.)

Non-Microblogging Platforms

3rd Generation Definitions	Facebook	Instagram	TikTok	LinkedIn	Snapchat	Pinterest
1 Prioritise users' free expression and access to information ...	2 *	1 *	1 *	0 *	1	1 *
2 Prioritise and set as a continuing goal the improvement of society ...	1	0 *	0	0	0	1
3 Clearly and thoroughly explain what speech is not permissible ...	2 *	2	1 *	2	1 *	2
4 Reject the attention economy and personalised algorithmic ...	-1	-1	-1	1	-1	0
5 Address the United Nations Bill of Human Rights and detail how ...	0	0	0	0	0	0
6 Address corporate social responsibility ...	3	3	0	3	3	3
7 Require only the necessary licenses for user-supplied data ...	-1	0	-1	3	1 *	0 *
8 Provide complete transparency regarding published policies ...	2	2	1	2	2	1
9 Publish detailed transparency reports on content removals ...	2 *	2 *	2 *	2 *	2 *	2 *
10 Publish regularly updated statistics on data relating to users ...	1 *	0	2	1	2	0
11 Publish as much data as possible without breaching privacy ...	2	0	1	2	2	0
12 Permit a yearly independent investigation and subsequent ...	1 *	1 *	0	0	0	0
13 Open source all code which has involvement in the ...	0 *	0	0	0	0	0
14 Establish and publicise a full disclosure on platform ownership...	1 *	1 *	0	2	2	1 *

Definitions Comparisons Chart (Cont.)

	3rd Generation Definitions	Facebook	Instagram	TikTok	LinkedIn	Snapchat	Pinterest
15	Openly make available to any self-identifying person over the age of...	0	0	0	0	0	0
16	Ensure an exceptionally high standard of data protection ...	0 *	2	0	2	2	2
17	Allow for and support fully anonymous users.	-1 *	1	0	-1	2	2
18	Actively and proactively prevent the exploitation of the platform to...	0 *	0	1 *	2	1	1
19	Review and verify the truthfulness of all political advertisements.	0 *	0	3 *	3 *	2 *	3 *
20	Actively provide preventive guidance for those searching for ...	3	3	3	2 *	1 *	3 *
21	Actively provide functionality to aid those with suspected social ...	1	1 *	2 *	1 *	3	0
22	Reject the use and support of tools that facilitate and aid the ...	-1 *	-1 *	-1	2	-1	1 *
23	Provide optional functionality to reduce or restrict access to types...	1 *	-1	2 *	2 *	1 *	0
24	Provide an efficient and timely avenue of appeal for users who ...	2 *	2	1 *	2 *	2	2
25	Facilitate and permit the existence of third-party clients in a ...	0 *	0	0	0	0	0
26	Promote healthy competition by interworking services with ...	0	0	0	0	0	0
27	Ensure continued value of investment in the event of ...	0	0	0	0	0	0
28	Expand the capacity, geographic, and linguistic diversity of content ...	3	3	3	3	3	3
29	Ensure advertisements cannot be tailored specifically to single ...	1 *	1 *	1 *	1 *	1 *	1 *

Definitions Comparisons Chart (Cont.)

3rd Generation Definitions	Facebook	Instagram	TikTok	LinkedIn	Snapchat	Pinterest
30 Make freely available to anyone self identifying as a researcher ...	0	0	0	0	0	0
31 Provide a warning of no less than 7 days to relevant parties about ...	0	2	1	3 *	2 *	2 *
32 Allow any user to close their account and delete all personal ...	0 *	3	2	3	3	1 *
33 Resist governmental pressure from any country on any issues that ...	1 *	1	0	1	1	1
Maximum Available Points 99	25	28	24	44	38	33

Background

Sonata was conceptualised following the creation of the definition of a third-generation social media platform. All definition components have been included within the initial blueprint, to be realised as part of the 1.0 release, expected in late 2023.

Sonata is being developed from the ground up to allow the whole vision to be accomplished without restriction.

Platform Stance and Direction

With the prevailing platforms relenting to pressure from advertisers influencing their content policies and new platforms generally taking a rigid stance on prohibiting sensitive content, Sonata's stance is to commit as heavily to freedom of expression as is possible.

This will be achieved through a system called 'amplification', which will allow for the platform-level promotion of users who follow content guidelines. This system enables Sonata to distinguish between two content policies, namely 'Prohibited Content Policy', which will contain illegal material such as stolen personal bank information, and 'Amplification Policy', which will outline all grey areas of speech that are legal but widely considered dangerous, harmful or immoral. Users who expressly go against the Prohibited Content guidelines will be necessarily excluded from the platform. In contrast, users who go against the Amplification Policy guidelines will see their platform-level support and promotion removed, but their presence and content will remain.

With Sonata allowing content that most other platforms would prohibit, this introduces a social duty of care over users who may not wish to view such material, particularly vulnerable users. The amplification system will ensure that access to such material will be a deliberate choice made by adult users.

In essence, freedom to express and freedom to ignore.

Justification

While there has recently been significant activity in the development of new platforms, none have yet attempted the significant ideological leap into embracing and prioritising social responsibility. Sonata is foundationally built on the comprising principles and is ready to insert itself as a leading example of what people should come to expect from social media.

A recent wave of interest and development has brought decentralised platforms into the spotlight as possible candidates for the future of social media. Sonata does not follow this direction, as decentralised platforms have inherent restrictions that obstruct the ability to prioritise social responsibility fully.

To expand on the limits of decentralised platforms:

It is likely outside of current practical technical ability to fully synchronise content beyond a certain scaling limit on a decentralised network, making such a service unsuitable for large-scale operations. Content processing and delivery at scale requires significant investment and expertise, leaving most current offerings to avoid full content synchronisation.

If not synchronising content, decentralised platforms:

- Isolate users into self-imposed echo chambers.
- Tend towards the concentration of users into single instances, which must grow the scope of their operation to handle the increased user base, developing by necessity a centralised, responsible entity and reducing the significance of the foundational decentralised nature.
- Exclude themselves from social responsibility by requiring all instances to take responsibility for the content on that instance.
- Have instances run by unknown people, which generates privacy and safety concerns.
- Have instances which are too small in scale to justify paid moderation, leaving the work to volunteers or the owner, who may over-moderate and restrict expression or under-moderate and facilitate abuse.
- Have instances that are likely to vanish without notice, especially if it costs more than the owner can manage.
- Have instances which restrict sign-ups due to the burden of moderation and server costs.

Intention

Sonata is intended to be the new standard in microblogging, providing free and open freedom of expression and information. The initial version 1.0 release will focus on basic functionality with particular attention to wholly encompassing and complete, ideologically driven policies which set the stage for later feature expansion. Feature plans up to version 1.5 will focus on improving the general experience and ability to express oneself through microblogging.

Sonata sets out the following goals:

- To earn a reputation for being the global platform most supportive of freedom of expression.
- To earn a reputation for being the most socially responsible global and large-scale platform.
- To continually innovate and provide new products and features which enhance societal discourse, expression, progression and well-being.
- To remain financially self-sufficient without selling any portion of the company's share capital and without restricting the base functionality of posting and viewing posts through payment requirements.
- To stay genuinely neutral in content moderation by abiding by and only taking actions expressly permitted by a clearly defined set of policies.
- To respect every user's human rights in any situation.
- Bring about positive and constructive tools and services to aid and further the evolution of social media platforms for the good of society.

Sonata Policies

Further to the articles defining a platform as third-generation, Sonata will maintain a set of internal company policies[15].

Motto : 'For the betterment of society'.

Policy: Stand up for our company values

When making decisions, always side with the established company values, being:

- Support truth
- Honesty in all things
- Respect to all users
- Pursue innovation
- Protect authentic human expression
- Trust is earned.

Policy: Strive for positive impact.

The primary consideration in decision making and the direction of the company will be to create and promote a positive impact on society.

Policy: Avoid decisions that will lead to a profit & growth driven future.

A profit and growth driven social media company will find it more difficult to make decisions that improve society, if they also impact profit or potential profit. The following decisions should be avoided:

- Converting to a publicly traded company
- Selling equity to private investors
- Committing to new, expensive to run features before suitable reserves of funds have been built.
- Scrapping otherwise beneficial features which do not contribute to revenue.

Sonata Policies

Policy: Seek alternative sources of revenue to ads while ensuring a fully free service at the point of use.

Ads bring one major benefit – allowing for platforms to be free at the point of use. However they are often unwanted and intrusive and often slowly start to cause to policy changes as a platform seeks to further monetize their users.

Seek out new ways to generate income that don't have these drawbacks, either through the platform directly, within sub platform level services, or within entirely separate services that leverage the platform's status, trust and influence.

Policy: Fairness for all involved.

All staff, contractors, affiliates and providers should be treated as valuable elements of the company and treated appropriately.

Policy: Freedom of expression and freedom to ignore

Users should be free to speak their mind as much as can be legally entertained. They should not live in fear of being banned by arbitrarily governed moderation.

Additionally, users should not be forced to be subject to the expression of others.

Policy: Protect the planet.

Our planet is our most precious resource, decisions should be made with this in mind.

A push towards fully net zero power sources should be found as soon as economically viable.

Executive travel should be restricted as much as possible while conducting business.

Sonata Policies

Policy: Encourage critical thinking

Social media platforms have made it too easy to take on the opinions of others with little oversight, hindering the development of critically analysing topics or opinions. Strive to counter this as a feature of the platform.

Policy: Mandate through policy

All mandates the company's teams operate through should be based in policy doctrines and any decisions taken should be justifiable by the established policies.

Policy: Transparency Throughout

Any data that can be made public, in that it does not compromise the privacy of a user or put the operation of the company at any risks, should be made public.

Policy: Political neutrality

- Do not encourage or promote or discourage or demote any people or material due to their support or opposition of an active political party.
- Do not allow any political party to spend more than 50% of the total published budget of their major competitors on advertising. Major competitors is defined as any party that has within the past 10 years received more than 25% of the vote in any major election.
- Do not allow targeted political ads beyond country and county.

Policy: Work towards fully encompassing the full definition of a 3rd generation social network. When achieved, maintain and improve on the principles.

Launching with the full principles of a 3rd generation social network fully realised will not be possible. These should be worked towards as a priority, with an estimated date set.

Sonata Policies

Policy: Take drastic measures where required to keep company values intact.

Unacceptable and ill thought out changes to law in the countries Sonata operates in may require relocation of servers, staff or the company itself.

Policy: Protect all users from outside parties

No user should suffer exposure of their privacy or data from outside sources seeking to obtain it through back channels.

Private requests for non public information about any individual user sent to Sonata for any reason and from any source must be rejected.

Exceptions:

- When legally forced, through laws or through motions of the courts of Iceland
- Data concerning individuals banned from Sonata for posting schedule 1 or 2 prohibited content may be made available to law enforcement.

Policy: Publicly support and pursue company policies externally

While remaining politically neutral, Sonata should actively support and promote our company polices by:

- Defending and promoting free expression
- Promoting and furthering the uptake of social responsibility by exposing and reporting on the exploitation of users by social media companies.
- Promoting the concept of a third generation social media platform and all of the definitions which comprise it.
- Promoting critical thinking.
- Supporting and promoting people and institutions who aim to make or inspire improvements which relate to our policies.

Sonata Full Description at Version 1.0

Sonata v1.0 release expected Q4 2023

Summary

Sonata is a free online microblogging solution provided through a web browser and native Android and iOS apps. All aspects defining a 3rd generation social media company have been realised. Due to their progressive stance towards online expression, the company (Sonata Social ehf) which owns and runs Sonata is incorporated in Iceland.

Sonata is named as an acronym: **SO**cial **N**etwork **A**chieved **T**hrough **A**mplification.

Users may create unlimited posts, called 'notes', within a 440-character limit[16]. Longer messages may be made by forming threads, which will be natively displayed in an easily readable fashion. All notes may have one image attached, or one active hyperlink, which will cause a preview to be shown along with the note.

All content on Sonata will be accessible without requiring a login.

Users may follow other users to see their posts automatically appear in their homepage feed

Content Policy

All content will be permitted on Sonata unless there is a significant pre-defined reason that it must be removed. No moderation action may be taken unless expressly permitted within the content policies.

This creates the need for multiple classifications of content:

1. Prohibited content: as described by the Prohibited Content Policy.
2. Unsupported content: as described by the Amplification Policy.
3. Sensitive content: as described by the Sensitive Content & Flag Policy.
4. Supported content: defined as any content falling outside of the previous classifications.

Prohibited Content

Relevant policy: Prohibited Content Policy

The content described within this policy contains any material which makes the continuation of Sonata as a service impossible. This includes illegal content such as CSAM, material that gives extreme risk for harm, copyrighted material, revenge porn, and technologically abusive behaviour, such as automated spam and attempts to damage the Sonata infrastructure.

The Prohibited Content Policy is split into several schedules, each conveying the appropriate action to be taken according to the severity of the material posted. Some schedules call for the immediate removal of the content and the associated account, whereas lower schedules call only for content removal and warnings.

Unsupported Content

Relevant policy: Amplification Policy

The content described within this policy is material that would typically be disallowed on other platforms. This includes hateful material and material with a high potential to cause non physical harm to another through instruction or influence.

Like the Prohibited Content Policy, the Amplification Policy is split into multiple schedules in order to mandate appropriate moderation action. Repeat infractions may see escalation to a higher schedule.

Sensitive Content

Relevant policy: Sensitive Content & Flag Policy.
Sensitive content is defined as content that may not be suitable for all audiences. This could be through personal choice or through policy, such as not showing sexual content to a person below the age of 18.

Significant overlap exists between sensitive content and content described in the Amplification Policy. During the use of Sonata, users will sometimes be presented with a warning before accessing content flagged as sensitive.

Sensitive Content (Cont.)

The flags which may be used are:

- Nudity (Non-sexual)
- Sexual content (including sexualised nudity)
- Violence
- Sensitive (such as content likely to be shocking to most users)
- Warning (material reported by users but not yet verified by a Sonata moderator)
- Bot (content created by an automated system)

Supported Content

Relevant policy: Amplification Policy

Any material not addressed by other classifications will automatically be supported. This will be handled by the amplification system, which will allow Sonata to assign confidence in a user's genuine expression by the content they produce along with other signs, such as their activity levels, verification level and from an endorsement by other users.

Higher amplification levels will allow Sonata to promote a user and their content freely, according to the Content and User Recommendation Policy.

Privacy Overview

All data held by Sonata on an individual user must be fully described, including information on how it is stored, for how long, who it is shared with and a full justification of why it is needed.

Notably, Sonata does not store email addresses in a readable format, instead storing them in the same encrypted way as passwords. This allows the system to verify users logging in, to email for forgotten 24 password requests, and to email at the point of user login, but not send emails to the user at any other time, making Sonata immune to mass exposure of personal emails due to unauthorised inner system access or hacking.

As per EU regulations, the ability to reject targeted advertising is given separately from the general terms of service. A user may accept the terms of service but reject any targeted advertising. This decision may be changed at any time and does not impact the service offered to the user outside of a change in advertisements shown.

The full breakdown of data kept on each user, with all information mentioned above, is referenced in the Privacy Policy.

Protections

With the content policies as open as possible, Sonata needs to maintain a dedication to the protection of users that either require it or request it, as needed, to maintain a total commitment to social responsibility.

Types of Protections

1. Exposure To Prohibited Content
2. Exposure To Harmful Content
3. Harassment From Other Users
4. Manipulation Or Contact With Malicious Intent From Other Users
5. Harassment From Outside Actors

Exposure To Prohibited Content

Users will expect a level of safety while using the platform, that they will not be exposed to illegal, dangerous and shocking material. Sonata's content detection, moderation and user reports will look to prevent any such exposure from occurring.

Exposure To Harmful Content

As with prohibited content, users expect not to stumble across harmful content. However, some users will intentionally seek out harmful content which remains available on the platform. Sonata carries a duty of care to ensure that these users are not promoted such content and do not necessarily have an easy time finding and accessing it. Some users will find that they are required to read the information provided by Sonata on the topic, including external links and contact information of institutions that offer relevant help. Some users may also be required to watch short educational videos provided by Sonata on the topic before accessing the content to help teach them the potential harm they are exposing themselves to.

Harassment From Other Users

Targeted harassment1[17], whereby a user is continually responded to and or mentioned by a group of people beyond what would be expected from their content's reach; or from multiple repeated instances of contact from the same users, may constitute harassment. Users falling victim will benefit from automatic detection of this activity, resulting in junk notification filtering.

Users may also enable a 'safe mode', isolating them from outside users and de-amplifing themselves. Users may report that they are being harassed, triggering a higher level of automatic detection.

Moderation action may also be taken, leading to de-amplification and admin-triggered mutes.

Conversely, it is essential to distinguish between harassment and simple disagreement; someone with exceptional opinions will naturally find a lot of natural backlash and rebuttal to their posts, which should not be considered harassment.

Manipulation Or Communication With Malicious Intent From Other Users

Some users may attempt to mislead others, particularly those most vulnerable. Sonata employs methods to detect known forms of such behaviour to help safeguard these vulnerable users. The resulting actions will include mutes, bans and de-amplification, as described and defined in the Content Policy.

Harassment From Outside Actors

It is commonplace in certain countries for whistle-blowers, journalists and dissenting citizens to be subject to an inquiry from their governments. Sonata will stand up to these inquiries as long as these requests are not made from Sonata's country of origin, Iceland. Iceland was chosen as a host country partially due to the unlikelihood of this occurring. Sonata will maintain a policy of permitting anonymity for users who require it.

In response to private companies seeking data about specific individuals, Sonata will defend the rights of its users, refusing to give any data that is not already public. If necessary, the same measure mentioned above will be carried out.

Sonata maintains a canary[18], one for each county, for each of the following:

- Requests from official government institutions for data on certain user(s) within their jurisdiction.
- Request from official government institution for data on certain user(s) outside their jurisdiction.
- Fulfilled null value[19] request for official government institution for data on certain user(s).
- Fulfilled full or partial data requests for official government institutions for data on certain user(s).

Note the exception to these canaries, data pertaining to users that have violated section 1 or section 2 of the prohibited content policy, which covers disseminating child abuse material, bomb-making guides and other material of similar gravity.

Please see the Content Policy for the full details.

Protections in Practice

A definition can be made for multiple groups of people that might require protection on the platform:

1. Children (aged 13+)
2. Vulnerable Adults
3. Adults Requesting Protections

Children

People under the age of 18 must not be exposed to sexual nudity or any other such material of a sexual nature, nor violent content or sensitive content, which may include material that has a reasonable possibility to cause harm.

Sonata for children is a substantially different experience than it is for adults, with a significant amount of content restricted from view where it has been flagged as in any way inappropriate.

Adult accounts will not trigger notifications on children's accounts.

Grooming detection AI will scan any communication between adult and child accounts. In the future, a further restriction of communication between these two groups may be added, especially in coordination with a future release of direct messaging functionality.

Vulnerable Adults

Adults detected searching for certain known harmful material may be designated as vulnerable and subject to helpful material, either as a dismissible suggestion or by a forceful completion of tasks, such as entirely watching an educational video, as is determined as appropriate[20].

Adults Requesting Protections

Users may request that any types of content detailed within the Sensitive Content & Flag Policy be removed from their feeds and suggestions.

Protection Related Challenges

Protection and moderation is often the most challenging part of hosting and maintaining a large-scale platform. It is expected that managing this will constitute a large percentage of the Sonata team's overall efforts.

Continued development into crowd-sourced information regarding content flagging will likely be required, as well as investment into AI systems to work in tandem with human moderators to implement the Content Policy accurately.

Moderation

Sonata employs a system of administrator moderation, user-sourced reporting and automatic assessments through keyword evaluation and AI evaluation to effectively monitor the content published by its users.

Users are also incentivised to self-moderate their posts through the flagging system. Failure to flag sensitive content appropriately results in temporary deamplification as described in the Amplification Policy.

State Imposed Censorship

While it would be preferred that this section were blank, some countries that Sonata will be available within have content removal laws that require content to be removed following a report, which may undermine our amplification system.

Sonata will respond to these in the following ways:

- Any content that falls under the prohibited content definition will be removed.
- Any content which falls under the unsupported content definition will be made unavailable to users of the complaining country.
- Any content that falls into sensitive or supported content definitions will be defended. To avoid blocking or fines, a decision may be made to make this content unavailable in the complaining country

Sonata may also ignore these requests, knowing that this carries the risk of the platform being blocked in that country.

All requests, whether carried out or not, will be published in yearly transparency reports. Relevant country canaries will be removed.

User Types

As Sonata is a platform prioritising freedom of expression, users must find it safe to assume that all other users are genuine people providing their genuine expression. Moderation efforts will look to remove users that work against this.

The following user types can be identified:

- **Genuine User:** A human[21] expressing themselves. This can be an identifiable person, an anonymous person or a character. A natural person may control many such accounts. In later versions of Sonata, API control may be used, particularly for post-scheduling.
- **Automated User:** A bot that posts content that a human did not write or is a relay of information published elsewhere. It may control only a particular channel of a user's account. It must be flagged as 'bot' content.
- **Business, Institution or entity:** An account controlled by one or more people that reflects the expression of the entity.
- **Bot Farm User:** An account of a fake person, controlled as part of many accounts by a small group of operators. When discovered, users such as these will be removed, along with all users that are part of the bot farm.
- **AI User:** A user posting content generated by an AI[22] that is presented in an attempt to pass off as genuine human expression, whether operated by an automated process or through human control. AI Users are not permitted on Sonata and will face immediate removal.
- **Human Operated Troll Farm User:** Often state-sponsored, these are groups of users operated by paid humans to attempt to manipulate opinions. When found, these users will be removed.

Some genuine users may be assigned a subtype, which helps to identify them on Sonata further.

The following genuine user subtypes can also be identified:

- **Journalist:** Special accounts that are immune from automated content moderation.[23]
- **Parody:** Intentional impersonations that do not genuinely pretend to be the person or character they portray.
- **Memorial:** When a person is deceased, their account can either be retired or converted into a memorial subtype, which permits the owners of the person's estate to continue the account in their memory.
- **Unofficial:** Used for fan accounts and clubs that do not identify as parodies.

Platform Features

Sonata presents a number of features beyond that of a standard microblogging platform, allowing it to stand out from other services. There are also many more features in development, as described within the roadmap.

Notes

Each post created on Sonata is referred to as a 'note', which can contain up to 440 characters of text.

Users may post unlimited notes.

Notes can contain an image uploaded by the user. They may also include a link, which may cause a preview of that page to show along with the note.

Notes can be edited, including modifying links and images. Other users may 'renote' a user's note, displaying it on their profile and adding it to their feeds. They may add their text and create a quote, which shows above the quoted note's text. Notes that are subsequently edited after a renote will not cause the renote's text to change.

Users may engage with another user's notes by 'liking', or replying. A reply consists of a new note that is permanently applied to the parent note and follows the same rules on text length

Channels

Every note created on Sonata must be sorted into user-defined 'channels', which are then used for grouping into topics. Users may create and edit these channels at any time.

Channels allow Sonata users to freely express themselves without worrying that their posts are of varying quality or cover a wide range of topics.

Other users can follow their own choice of user channels, giving them granular control over which topics they'd like to see in their feeds from each person.

Commonly tagged channels allow globally available feeds, showing content from all channels of the same type.

Amplification

Users who utilise Sonata will gradually gain amplification levels, with one being granted for each milestone[24] they achieve. At level 3, this reaches its full effectiveness, allowing Sonata to confidently promote that user and their notes across the platform.

Users can see their amplification level on their settings page, along with details on how they achieved each level and how they can earn more. It is not possible to see another user's amplification level.

As amplification is an impartial measure of how genuine and trustable a user is, all users must have the potential to reach the highest level. While there will be ways to advance by paying – as this evidences that a user is not fake, two rules must be adhered to:

1. It shall not be possible to use money as the sole means to reach the highest level of effectiveness.
2. It shall be possible and within the means of all users to reach the highest level of effectiveness without requiring them to pay.

Violating the amplification part of the Content Policy will cause the user's level to drop to at least -1. All positive milestones will be ignored until the violation expires.

User Links & Verification

Users may display links to their accounts on other services directly on their profile. The user adds and controls these and may provide links to any location.

Certain service links may also benefit from Sonata's verification system. With the spread of so many services being widely used today, Sonata aims to be a central point of reference for a user's outlets, though this requires the issue of falsehoods giving way to scams to be addressed. Sonata tackles this by providing automatic verification by ensuring that an account or page belongs to the person who linked to it if this is claimed.

User Links & Verification (Cont.)

Sonata provides automated verification for the following:

- Twitter
- Youtube
- Instagram
- Website (via TXT DNS or homepage meta tag)

Automatic verification for more services is in development.

Announcements

To prevent the need for algorithmic notifications, which attempt to figure out what's important enough to promote, Sonata provides a more direct solution that gives control to users and their followers.

Users may elevate one of their new posts every 24 hours to become an announcement. This will subsequently cause a notification to be issued to each following user.

When following others, users can choose whether also to follow their announcements. This is provided at both account and channel levels.

The announcement system lets users be confident that their most important messages will not be missed. It is also somewhat self-regulating – if users abuse it by launching announcements over trivial matters, their followers can block further announcements from them.

User Discovery and Content Exploration

As Sonata rejects the attention economy, displaying content based on engagement is not permitted. Therefore, a more ethical system of recommendation must be adopted.

For recommending users to follow, Sonata will, by default, first filter out all users with lower than level 2 amplification (users may choose to set this between levels 0 and 3). It will then show a random selection of users and some of their recent content and recently used channels. Several non-algorithmic factors will influence this list, including geographical distance, recent activity and amplification level.

User Discovery and Content Exploration (Cont.)

Content exploration is presented in multiple forms:

- Location-based feeds show in chronological order posts from any user with higher than amplification level 1, with feeds for the nearby area, the same country and one for notes made globally.
- Trending hashtags give an insight into current hot topics, and every hashtag is a feed showing chronological posts that contain that hashtag.
- Channel feeds show channel names that are commonly used on Sonata, with their associated notes showing as part of the feed in chronological order.
- Journalist feeds show all content posted by verified journalists.

Verification

Knowing whether a notable person's account is genuinely owned by that person is a common issue on all platforms. Another problem is knowing whether any person is actually who they say they are. Sonata solves both of these issues by providing multiple levels of verification.

Notable people or organisations (as described in the Verification Policy) will receive a full-colour tick mark next to their name after application. This provides them with one amplification level and no other further benefits regarding content recommendation.

Any user may apply for verification of their name, location or work, which will display as a tick with a transparent background. Due to the costs involved with administrating this, this may need to be a paid feature.

Technical Overview

Sonata was designed and built from the ground to be scalable to meet the growing requirements of a global platform. It was also designed to prevent costs from becoming unsustainable by developing certain in-house systems rather than using ready-made cloud solutions.

Client Servers

Role: Providing endpoints and dealing with most connections between users and Sonata.

Client Servers (Cont.)

Setup: These servers use Open Swoole, a stateful PHP-based system that improves performance dramatically over plain PHP. This also makes for an effective socket-based client/server communication implementation.

Media Servers

Role: Processing and providing images.

Setup: Rather than using expensive cloud-based providers, Sonata uses a custom-designed system to store, process and back up images. An estimated 10-15x cost reduction can be achieved over using cloud services.

Mobile Applications

Role: Providing an interface for Sonata for the majority of users on both iOS and Android

Setup: Coded using Flutter, a modern language designed to simplify the process of producing apps for both operating systems

Desktop & Web Mobile

Role: Providing an interface for Sonata for computer users and phone users unable or unwilling to use the app.

Setup: Coded using React, our choice of JavaScript frameworks to provide the full Sonata functionality.

Database

Role: Store all data created and used by Sonata, such as user information and note contents.

Setup: MYSQL using Vitess, an extension developed and used by Youtube internally.

Accessory Systems

Role: Performing tasks which aid the general function of the platform.

Accessory Systems (Cont.)

Setup: Multiple servers performing tasks such as feed generation, external resource crawling, administration and moderation.

Caching

Role: Storing regularly used data to reduce the load on the database

Setup: Multiple servers storing and providing data for internal use.

Roadmap Including Timelines

v1.0 - Core Platform (Closed Alpha) *- Launched*

- Base microblog features and general platform function
- Static Informational Site
- Whitepaper
- Threads
- Profile Links
- Data Insights (Privacy)
- Channels
- Simple distinction of under 18 accounts
- Chronological Feed
- Profiles
- Discord

v1.1 - Refine Core Features (Closed Beta)

- Dark Mode
- Upload image tools (Crop, Resize)
- Mentions
- Hashtags
- Account types
- Keyword / AI safety scanning
- Reporting content
- Saved notes
- Iceland Company Fully Established
- More note and profile settings

v.1.2 - (Open Beta)

- Location based feeds
- Image safety scanning
- Shareable link list (Standalone)
- Standalone threads (Threadreader)
- Trending hashtags
- Amplification
- Paid accounts to support development and limit reliance on ads
- Auto external page verification
- Announcements

v1.3 - 3rd Generation Definition Reached

- AI Aided Moderation
- Canaries
- Verified Users - Notable Persons
- Educational Resources on certain topics
- Simple tools to allow users to easily take a break
- Safe mode
- Appeals - for moderated content

v1.4

- Business Update
- Anchor Tags (&)
- Sonata Echo (sync to Mastodon, Twitter, others)
- Child protection update
- Affiliations
- Platform Translation - 12 languages
- Conversation view for public discussions. Upvotes & downvotes for replies.
- Verified Users - Everyone
- Direct Journalist support
- Create by email
- Multiple images per note

v1.5

- Tools for users with a high number of followers
- API
- (if sufficient users & possible financially) First party ads only
- Sonata MyData creation and integration
- Universal tags update – Community
- Site quote
- Circles (group system)
- Review all external tools (analytics, fonts) to see if they can be removed
- User Note Translation
- External website support
- DMs
- Platform Translation - 24 languages
- Multi user support

v1.6

- Comment Anywhere
- Analytics
- Dynamic Notes
- Release platform metadata publicly
- Important channels - don't miss an update
- Web3 login and account creation
- Platform Translation - 48 languages

v1.7

- Social Trust
- Social Tagging
- Personal achievement list & verification
- Ad platform for select other microblogs
- Custom feeds
- Platform Translation - 96 languages

v1.8

- Platform Translation - all required languages

v1.9

- File attachment to notes

Future Plans

The launch of Sonata is the start of an expansive overall plan, with the current roadmap leading up to the second stage of the platform. By this time, the aim is to have a significant global user base, stabilised finances, and to have substantially improved people's ability to express themselves worldwide.

From here, expansion will be made outwards, creating new services and providing new possibilities on the back of the Sonata platform. Exploration will be made for alternative ways for Sonata to earn enough to continue its operations. All current social media platforms are almost entirely ad-supported while remaining free at the point of use. This leads to pressure from ad purchasers to maintain increasingly restrictive content policies. Sonata aims to rid itself of this pressure while always remaining free at the point of use by creating related services, which will feed its profits into Sonata's revenues. Any riskier ventures will be created as independent companies, removing any risk to Sonata.

This section overviews some of these planned services and their intentions. All service names are likely to change.

MyData

Structure: Open Source, App

Finances: Initially funded by Sonata, fund may be established for project continuity.

Monetisation Potential: Minimal.

Profit Expectations: Minimal. Any surplus will be kept in project as an operational reserve.

An open-source system for users to store their data on their own device, rather than entrusting it with any company. Companies will provide optional relays to enable constant availability.

Content Safety

Structure: Private API Service

Finances: Funded by Sonata

Monetisation Potential: Moderate. Requires many services that require assistance and can afford to pay for it.

Profit Expectations: Some, to be added to Sonata revenues.

Content Safety (Cont.)

An expansion to the Sonata internal content safety system, Content Safety will provide enterprise-level protection to other user-generated content platforms, such as Mastodon instances and startup social media platforms. Providing a service which allows user-generated content is quickly becoming more complex in many countries due to excessive liability placed upon the platform to moderate, which leads to over-moderation and, as a result, limits expression. Sonata will help by labelling illegal content appropriately, allowing the platform to take action.

Verification

Structure: Private API Service

Finances: Funded by Sonata

Monetisation Potential: Moderate. Requires many services but may be of interest to a wider range of companies than Content Safety is.

Profit Expectations: Some, to be added to Sonata revenues.

An expansion to the Sonata verification system, Verification will verify a person or a business for use on another platform. This will allow small services to have verified users backed by an authoritative platform.

Follow Hub

Structure: Open Source, App

Finances: Initially funded by Sonata, fund may be established for project continuity.

Monetisation Potential: Minimal.

Profit Expectations: Minimal.

A system which will help users keep track of who they follow on which platform. Intended especially as an aid for new platforms, for those compatible, it will allow users to import followers from other platforms.

As a secondary function, it will allow followed users to create a request to follow on a new or existing platform, aiding with the transfer to a new platform.

Permissions will be made available to allow users to control or block these requests.

Responsible Social Media

Structure: Council

Finances: Funded by Sonata, other social media platforms.

Monetisation Potential: None.

Profit Expectations: None.

An industry self-regulator that aims to create standards for all other platforms to follow. It will keep constant reviews of all major platforms, finding ways that improvements can be made in respect of social responsibility. Only 3rd Generation platforms will be permitted as council representatives, while all platforms with more than 100,000 users will be allowed to join as members.

Governance and Ownership

To keep as much creative control over Sonata as possible, the aim is to grow the platform as much as possible without investment. Keeping Sonata privately owned will ensure the safety of its policies, away from the pressure of investors requiring growth and profit.

Sonata is owned and operated by Matt Beck, a software engineer and tech entrepreneur from Brighton, England.

Beck has created and operated platforms that facilitate and promote user-generated content since 2014.

Before the 1.0 launch, a company will be established in Iceland which will own and manage Sonata. A company has already been established in the UK. It is set to handle Sonata's UK-based operations, such as employing staff, server hosting and financial processing, as many payment processors still need to offer services to Iceland.

Development of This Whitepaper

The concept of social responsibility defining 3rd generation social media was conceived in October 2022 by Matt Beck and was developed over six months. Following this, feedback was sought from established human rights advocates, paving the way for the published definition within this document.

The concept for Sonata was developed by Beck during this time, including the complete blueprint of the platform, the technical architecture and the operational development plan.

Resources, Access & Contact

Contact in regards to this whitepaper at whitepaper@sonata.social

Contact the Sonata project at help@sonata.social

Contact Matt Beck directly at matthew@sonata.social

Appendix

This whitepaper references 'harm', which is essential to define precisely. If not otherwise expanded upon, 'harm' in this whitepaper is defined as a measurably negative impact upon a person which causes the impairment of any of the following capabilities:

- Life (suicide, self-harm by standard definition)
- Bodily health (lifestyle impairment including unhealthy sleep quality, drug, tobacco or alcohol intake, poor diet and lack of exercise)
- Senses, imagination and thought (ability to reason and critically think, ability to engage in activities which require concentrated thought)
- Emotions (depression, low self-esteem, social anxiety and poor interpersonal relationships)
- Practical reason (dissatisfaction with their life's progression, reduced ability to self direct)
- Affiliation (social isolation and loneliness)
- Play (sufficient reduction in free time to enjoy healthy recreational activities)

Adapted from *https://www.cambridge.org/core/journals/business-ethics-quarterly/article/ethics-of-theattention-economy-the-problem-of-social-mediaaddiction/1CC67609A12E9A912BB8A291FDFFE799*

Reference Notes & Sources

Whitepaper Notes & References

[1] Iceland specifically, as the host country of the service.

[2] https://www.cambridge.org/core/journals/business-ethics-quarterly/article/ethics-of-the-attention-economy-the_problem-of-social-mediaaddiction/1CC67609A12E9A912BB8A291FDFFE799 (Life: Several studies)

[3] Labelling content itself, rather than targeting the contributing user.

[4] https://counterhate.com/research/the-disinformation-dozen/

[5] Child sexual abuse material.

[6] As defined by: An individual seeking to record, collect and summarise fact checked information relating to current events for dissemination to the public.

[7] Almost all users of microblogging services use Twitter, with all other services combined boasting just a single digit percentage of Twitter's userbase.

[7] https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

[8] https://api.joinmastodon.org/statistics

[9] Each implementation of Mastodon, run by a separate group or individual is referred to as an instance.

[10] https://mastodonservers.net/servers

[11] https://www.businessinsider.com/twitter-competitor-hive-social-run-by-24-year-old-founder-2022-11?r=US&IR=T

[12] https://www.businessinsider.com/twitter-competitor-hive-social-run-by-24-year-old-founder-2022-11?r=US&IR=T

[13] https://www.thetimes.co.uk/article/from-the-flag-bearer-for-free-speech-to-scapegoat-parler-is-fighting-back-bmwcdfgf5

[14] https://www.eurekalert.org/news-releases/462763

[15] To be expanded upon following public feedback, in the run up to the version 1.3 launch.

[16] This is the longest character limit that was felt would allow the platform to retain its 'microblogging' description.

[17] Excessive interaction from one or multiple accounts towards a single person that goes significantly beyond simple response or rebuttals.

Reference Notes & Sources

[18] A statement made proactively, claiming that a service has not done something, such as complying with requests for information from a governmental agency. If this statement were to be made untrue the canary is removed, allowing the service to publicly reveal this fact without running afoul of any non disclosure laws or gagging orders. Similarly, it helps to alleviate suspicion from outside observers that an action was taken and subsequently prevented from being made publicly known.

[19] A response given which technically fulfils the request, but that contains no information which was not at any point publicly available. For example, if requested to provide all stored IP addresses used by a certain user, responding with an empty list (due to the information not being stored), rather than an outright refusal.

[20] As defined by Sonata policy.

[21] Also described as a natural person.

[22] Including generative AI such as GPT, LLaMa, Bard, PaLM, etc.

[23] However, prohibited content schedule 1 and 2 violations will lead to a suspension and review. Journalist accounts will be granted to users following an application.

[24] To be formally defined following initial tests and feedback. Currently, milestones include being an active user, avoiding amplification violations for a longer time, having a higher number of followers weighted by amplification level (in the region of 10,000), having an active membership and having a verification of any kind.

Reference Notes & Sources

3rd Generation Definition Notes & References

1. Sourced from Freedom House 2023. https://freedomhouse.org/report/freedom-net/2022/countering-authoritarian-overhaul-internet/policy-recommendations
2. Platform must state this, or something to follow the same meaning, prominently on their policy documents. Platform should be criticized and challenged with regards to policies, policy amendments and actions.
3. Sourced from Freedom House 2023. https://freedomhouse.org/report/freedom-net/2022/countering-authoritarian-overhaul-internet/policy-recommendations

9. Sourced from Freedom House 2023. https://freedomhouse.org/report/freedom-net/2022/countering-authoritarian-overhaul-internet/policy-recommendations

26. Social media services have always been very top heavy in the statistics covering the numbers of regular users. As more regulations set in and standards rise, the bars to forming new social media platforms are steadily rising, which may soon result in near impossibility for new platforms to get started without significant investment. Competition is typically a benefit to the end user, so this should be encouraged on a foundational level now, before the issue hits a critical point. Platforms should offer services such as advertising, moderation, content sharing, verification and the import and export of data in a standardised format.
27. Companies and enterprising individuals can spend significant amounts of money on increasing their following on social media platforms, which will become worthless if that platform collapses or recedes into obscurity. It should be made easy for a user to import their follows and followers (through invitation) on a different platform.

    Investment can be defined as time spent in building a following, or money spent achieving the same.
28. Sourced from Freedom House 2023. https://freedomhouse.org/report/freedom-net/2022/countering-authoritarian-overhaul-internet/policy-recommendations

Reference Notes & Sources

Comparison Chart Justifications

Twitter

1. No dedicated commitment to these specific things. Twitter also complies with governmental requests to block users or content even when this content does not go against Twitter's terms of service.
3. Translation of policies only cover around 20 languages, with Twitter's global userbase this should be sigificantly higher.
4. Twitter routinely distributes controversial material as part of its automated personalised content algorithms
7. Twitter requires a full, non revocable license of all content posted.
8. Twitter's policy documents are very short and vague.
9. Twitter's latest transparency report is now outdated (2021)
10. Unable to find this information published by Twitter directly.
11. In 2023, Twitter started charging very high fees for access to data.
13. In 2023, some Twitter code was leaked, but making this codebase open source was not commited to officially.
15. Twitter will provide generalised data to paying users.
19. Political advertisements are generally not permitted.
22. Filters are available on the Twitter app
24. Appeals are offered but not within set timeframes and not for all given reasons.
25. Third party client support was effectively removed in 2023
28. Twitter has many staff over the world, so 2 points are given in light of this

Reference Notes & Sources

Comparison Chart Justifications

Mastodon

1. Restrictive content policy for free expression.
3. Limited explanations of permitted content
7. No explicit licensing of content
8. Very limited detail on policies, many missing altogether
10. Not all data provided
14. Limited information given for some aspects
18. No proactive measures taken, except for lack of advertising.
22. None provided, but no dedication made to not providing them in future.
31. Does not currently add labels.
32. Not covered explicitly in any policy documents
33. Not covered explicitly in any official documents

Hive

1. Content not accessible without an account.
2. Attempting to make social media a better place.
3. Very limited breakdown available of permissible content
7. License page not available at time of check due to 404 (09 Aug 2023)
10. Some statistics given, but not regularly updated.
14. Owning company and persons named.
15. Not mentioned.
17. Requires name when signing up, but this is not verified
22. None provided, but no dedication made to not providing them in future.
23. Banned words and a NSFW filter is provided.
24. No appeal system advertised.

Reference Notes & Sources

Comparison Chart Justifications

Tribel

1. There are many reports of users facing harassment for expressing views that are not outwardly pro US Democrats. Sexual content is not permitted.
3. Limited breakdown of permitted speech
4. Uses self applied content labels to classify content and push to users.
10. Some statistics given.
14. Platform ownership and staff is disclosed.
19. On Tribel's privacy policy, advertisements are specifically mentioned to not be verified for truthfulness, which would also apply to political advertisements.
24. No information on appeals found.
25. No API provided
29. No relevant information found. Awarded 1 point due to difficulty of providing this.
31. Does not provide labels
32. According to Tribel's privacy policy, personal data is deleted on account deletion.

Threads

1. Requires an account to view content.
3. Content policy is somewhat vague.
7. Requires extreme licenses in regards to uploaded content, including 'modify, sub-license and create derivative works'
9. As threads is new, this may still be yet to come. This has been scored according to Instagram's record.
10. User totals and active users are regularly posted as part of marketing.
18. As threads is a new service, Meta's track record was used for this score, particularly their history with Cambridge Analytica
21. Has 'take a break' functionality built in
23. Ability to block offensive words, or custom words.
24. It is expected that this will be added in the near future.

Comparison Chart Justifications

Threads (Cont.)

26. Activitypub integration is promised but not yet implemented.
27. Activitypub integration is promised but not yet implemented.
31. Does not appear to currently add labels.
32. Currently not possible to remove Threads without removing Instagram as well.

Bluesky

3. Moderation is done through crowdsourcing
4. Provides an algorithm marketplace for content recommendation
7. Requires unfair licenses in regards to uploaded content, including 'modify'
10. User totals and active users are regularly posted as part of marketing.
13. Full open source not yet provided.
23. As the service is not fully launched at the time of writing, this cannot be fully checked

Facebook

1. Facebook removes significant amounts of content and has extensive 'do not post' content type lists. They do however provide dedicated support for journalists.
3. Extensive content policies are provided, however they should be improved with more specific examples.
9. Good level of transparency reports, but no indication of how machine learning is trained for use on content moderation.
10. Reports active and total users and ad impression counts.
12. As a public company, Meta releases reports on financial and legal matters.
13. Facebook does provide many open source projects, but none relating to these topics.
14. Platform ownership is disclosed. Data processing locations is somewhat disclosed, but includes 'amongst others'.
16. Facebook has suffered many data leaks.

Comparison Chart Justifications

Facebook (Cont.)

17. Facebook does not allow anonymous users and permits only 1 account per person.
18. Scored due to the Cambridge Analytica scandal
19. Regulates political ads separately. Provides disclaimers for ads that were not reviewed, but does not restrict unreviewed ads.
22. Tools available on phone app when taking a live photo.
23. Limited tools available.
24. Appeals provided but no dedication to respond or response times
25. All third party offerings are website wrappers
29. No commitment, but currently no evidence of this being possible.
32. Log in with Facebook' accounts will be lost if deleting a Facebook account.
33. Shows some resistance on issues like data requests and content moderation.

Instagram

1. Nudity not permitted.
2. Meta has published research showing that Instagram has a negative effect on teens mental health
9. Good level of transparency reports, but no indication of how machine learning is trained for use on content moderation.
12. As a public company, Meta releases reports on financial and legal matters.
14. Platform ownership is disclosed.
21. Has 'take a break' and daily time limit functionality built in
22. Tools available on phone app when taking a live photo.
29. No commitment, but currently no evidence of this being possible.

Reference Notes & Sources

Comparison Chart Justifications

TikTok

1. Nudity not permitted.
3. Limited list of permitted content provided
9. Good level of transparency reports, but no indication of how machine learning is trained for use on content moderation.
18. Does not publish data on advertisements presented, however they do not allow relations of politicians to advertise
19. Political advertisements not permitted
21. Blocks children from spending too much time on the app, has multiple tools to limit time spent
23. Includes video keywords
24. Provides account level appeals
29. No commitment, but currently no evidence of this being possible.

LinkedIn

1. Nudity not permitted, restrictive over content
9. Good level of transparency reports, but no indication of how machine learning is trained for use on content moderation.
19. Political advertisements are not permitted
20. Prevents discussion of dangerous topics, so guidance when searching is not required.
21. Provides optional automated detection of harmful content
23. Limited content permitted, so functionality not required.
24. No mention of how long it takes for appeals to be reviewed.
29. No commitment, but currently no evidence of this being possible.
31. Does not appear to add labels

Reference Notes & Sources

Comparison Chart Justifications

Snapchat

1. Nudity not permitted.
3. Permitted content lists are not very detailed
7. Includes transferable and sub-licensable
9. Good level of transparency reports, but no indication of how machine learning is trained for use on content moderation.
19. Provides a political ad library, requires approval on all political ads
20. Small banner shown to display help
23. Only content interests provided
29. No commitment, but currently no evidence of this being possible.
31. Does not provide labels

Pinterest

1. Nudity not permitted
7. Requires perpectual license over all content
9. Good level of transparency reports, but no indication of how machine learning is trained for use on content moderation.
14. Data storage and processing locations not shared.
19. Political advertisements are not permitted
20. Significant help offered.
22. Some filters provided, but not geared towards personal appearance.
29. No commitment, but currently no evidence of this being possible.
31. Does not provide labels
32. All user content is retained permanently

Changelog

Version 0.92.7 - *September 26, 2023*

- Added comparison chart for similar existing services

Version 0.92.8 - *August 19, 2024*

- Removed now defunct plans regarding external advertising