

Improving the News Feed on Hugging Face

Harsh Singhal

Executive Summary

Enhance Creator Discovery: By recommending similar creators and personalizing creator discovery, we can significantly boost audience engagement and offer more value. This approach makes it more likely for users to encounter content that resonates with them, fostering a more connected and active community.

Optimized Home Feed Rankings: By integrating diverse metrics such as popularity, social proof, and past click-through rates, we can enhance the relevance and appeal of the home feed. This optimization ensures that users are consistently engaged with the most pertinent and interesting content.

However, it's important to acknowledge potential risks in implementing these ideas.

For instance, basing home feed rankings primarily on popularity could inadvertently reduce visibility for new creators, contradicting our goal of fostering a diverse and inclusive platform.

To mitigate this risk, it's advisable to strike a balance that promotes both popular and emerging content.

I believe features like personalized user experience and enhanced creator discovery carry the lowest risk and offer substantial potential for positive impact. These could be initially tested with a sample group to gauge effectiveness and user response, ensuring that any implementation enhances the overall experience on Hugging Face.

Context

Hugging Face (HF) is the AI community where members collaborate on models, datasets and applications.

HF is also;

- A social platform for the AI community and a first of its kind.
- A platform that caters to developers (like Github) but is accessible to more stakeholders who may not necessarily be software developers.
- A cloud platform like AWS/GCP but only for specific types of tasks.

HF can also become;

- An education and influencer hub for the dissemination of content - tutorials style videos about models, paper discussions and influencer like videos about AI.
- AISuite - Equivalent of GSuite for ML/AI-first teams who wish to manage all development on one platform (private HF) and can choose to open-source artifacts selectively.

The following are key stakeholders in this community;

Stakeholder Persona	Why are they on HF?	What will delight them?
Company E.g., /microsoft /pytorch /ai4bharat /togethercomputer * BigTech doing open source * Only open source (Creators of libraries, models, datasets) * Data gatherers (CrowdFlower, Appen) * HITL providers (ScaleAI, NextWealth)	Share work directly with AI developers/researchers and users. Marketplace for Github (open-source), Mechanical Turk/ScaleAI (HITL)	Discovery by increasing followership. Social listening and analytics on what their followers are consuming.
Individual - Independent AI researcher/developer E.g., /TheBloke	Direct community participation and access to the consumers of their work.	Reactions to their content and followership. Discovering other independents to collaborate with.
Individual - Company/Org member E.g., /tnaumann	Develop authority in the AI social graph. Google Scholar++	Being discovered through an increase in followership. Potential to establish themselves as independents.
Individual - User E.g., /Harsh (me)	Gain knowledge and skills. Stay at the cutting edge of AI development and applications. Replace their LinkedIn/xAI feed for AI related content.	Receiving recommendations about relevant profiles to follow and personalized content to consume.

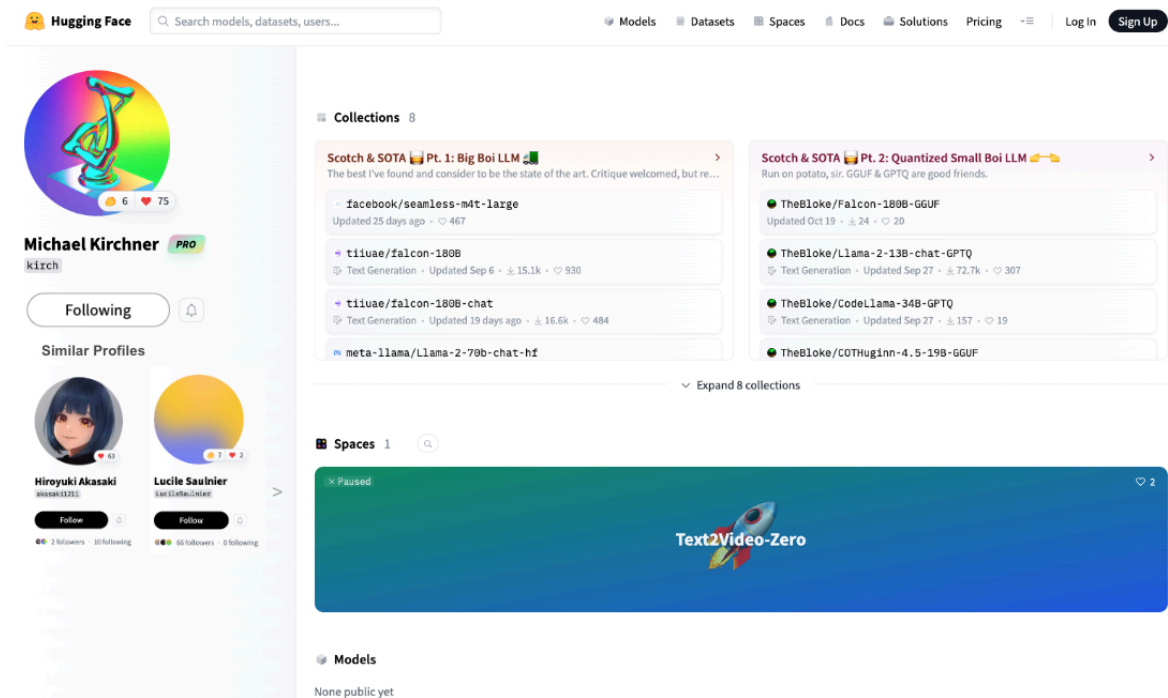
Key Thoughts

- Arguably there are more users consuming content on Hugging Face vs creators. This necessitates that the social features be improved.
- Social media features such as the following need to be developed;
 - **Home Feed Ranking** - content can be ranked and curated more effectively. Trending and suggested content mixed with following feed items in the Home feed.
 - **For You Feed** - discovery of content from entities not being followed by the user based on;
 - Algorithmic recommendations based on collaborative filtering
 - Content based on interest.
 - Content from creators similar to the ones followed.
 - **Similar Creators** - when I follow /TheBloke, who else should I be recommended based on;
 - Who else is typically followed along with /TheBloke by users.
 - Similar creators based on profile details and content they create.
 - **Creators Recommended For You** - based on all the entities I follow, my interests and interactions on HF, recommend a collection of entities I must follow.
 - **Personalized Notifications** - Personalized and relevant notifications to users.
 - **Topics** - Users might be interested in seeing collections of community discussions, Spaces, Models, Profiles collected under specific categories like LLM<=13B, Translation, Segmentation and so on.
 - Tasks are most analogous to Topics on HF today.
 - Topics can be shown to users during sign-up and can be followed by existing users to improve their Feed, Recommendations and Notifications.
 - **Search Ranking** - Search results should include social authority and activity.
 - $\text{Score} \sim x_1 \cdot \text{keyword relevance} + x_2 \cdot \text{followers} + x_3 \cdot \text{rxns} + x_4 \cdot \text{creations}$
where $x(i)$ are learned weights based on engagement data

The goal of this document is to showcase what product interventions Hugging Face should invest in to increase User engagement and increase the density of the AI community social graph.

Similar Creators

When a user follows a profile, show a Similar Profiles carousel below.



(Example mockup)

Bulk Follow

A bulk follow is a collection of creators grouped based on a specific topic. For e.g., we could group together the top model creators for Llama2-7b derivatives.

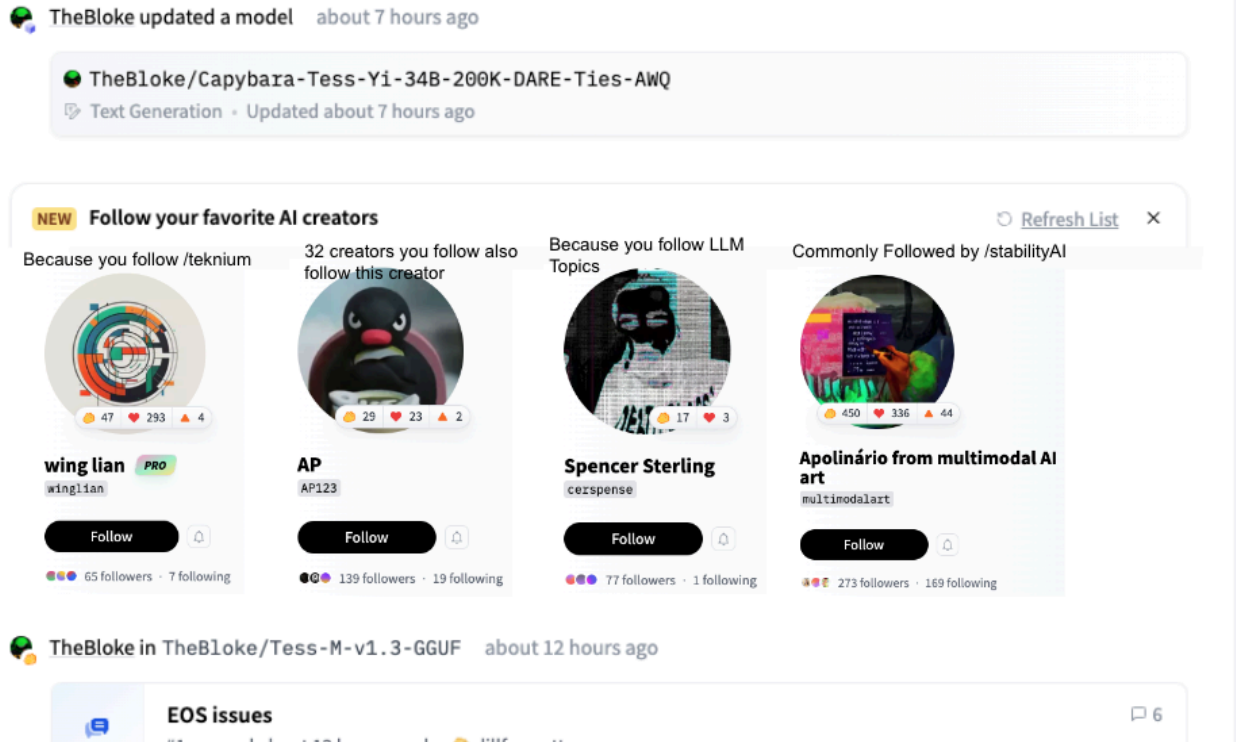
This group, when followed, will make the user follow all the creators in the group. Bulk Follow is a fast-track approach to get users to follow many relevant creators at once to jumpstart the flywheel for other social features.

Follow Your Favorite AI Creators (or Recommended For You)

The “Follow your favorite AI creators” carousel can be enhanced with the following interventions;

- Show mini profile cards of the recommended creators - this provides more context in terms of how popular is the creator (reactions and followers) and may nudge users to follow more popular profiles.
- Show justification text of why the creator is being recommended - this additional context leads to an increase in conversions and leads to an increase in profile clicks.

(Both these items are rendered below)



The following justification texts map directly to various signals.

Signal	Justification Text
Out of Network Affinity	<p>Recommend /teknium with the justification being Liked by /TheBloke</p> <p>I follow the /TheBloke and /TheBloke shares an affinity with /teknium which is why /teknium is being recommended to me.</p> <p>Here the affinity is if /TheBloke and /teknium are reacting to each other etc.</p>
Profile Views	<p>You visited their profile</p> <p>Recommend creators whose profile I visited often and recently.</p>
Friends and Friend of Friend	<p>From MSFT</p> <p>Recommend creators from the same organization.</p> <p>People you may know</p>

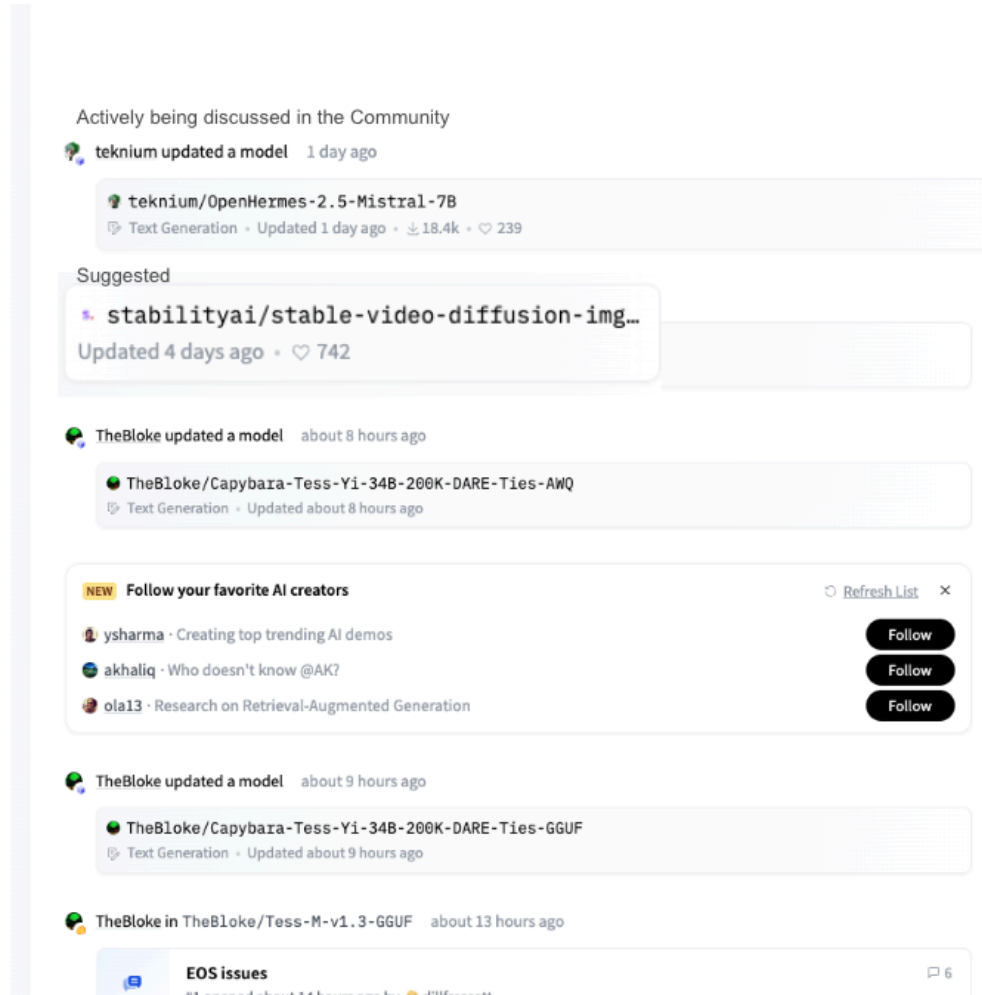
	Second degree connections from a different organization by connected through a trusted first-degree (in the same organization)
Contacts	<p>From your address book</p> <p>If your address book is uploaded to HF.</p>
ALS/Collab Filtering	<p>Popular Amongst People Like You</p> <p>These recommendations are the basis of a collaborative filtering approach between users and creators. The score between a user and a creator is basis - follow, reactions, profile visits and such.</p>
Top Topic Contributors	<p>Because you follow the Topic LLMs <= 7B</p> <p>Assuming Topics is a feature, creators who are top contributors of a Topic are recommended to users who follow said Topic.</p> <p>(Click through on this Topic will take the user to the ranked feed for LLMS <= 7B)</p>
Follow Similar Creators	<p>Recommend /multimodalart with the justification being Because you follow /stabilityAI</p> <p>I follow /stabilityAI and /multimodalart is similar to /stabilityAI</p> <p>Similarity between creators can be based on;</p> <ul style="list-style-type: none"> - Overlap of followers - Similar profiles - Similar content created
Commonly Followed With	<p>Commonly followed with /TheBloke</p> <p>I follow /TheBloke and I'm recommended /teknium because /TheBloke and /teknium share a large %age of their followers.</p> <p>(Basis Jaccardian similarity between the set of followers of /TheBloke and the set of followers of /teknium - this computation can be done by joining the edge list to itself and aggregating).</p>

Second Degree Connections	32 creators you follow, also follow /Undi Find creators that follow the creators you follow.
Follow of Follow	Followed by /TheBloke I follow /TheBloke. Show me the people that /TheBloke follows. (This is useful when I follow a top creator like /TheBloke. I would like to know who are creators that /TheBloke thought would be worth following).
Follow Back	They Follow You /teknium follows me and I don't yet follow /teknium. (Follow Back recommendations are based on the authority of the creator and are heavily curated before adding to the recommendations list)
Popular in your location	Popular in India /SanyamBhutani is popular amongst users who sign-up from India. Recommend /SanyamBhutani to others who are signing-up from India.

Home Feed Ranking

The following improvements can be made to the Home feed ranking;

- Algorithmic ranking based on
 - Popularity/Social proof - visits, likes and downloads and creator authority.
 - Relevance to the user basis interest, past click through rate.
 - Content from recently followed creators.
 - Content from tasks similar to the tasks the user has shown interest in.
 - Suggestions from Trending and Out of Network.
- Justification text to provide context and improve engagement with suggested content (see the text above the first two items in the screenshot below)



Signals to help source content to be included in the user's Home Feed along with the justification text.

Signal	Justification Text
Collaborative Filtering on content liked by users.	Suggested Based on your Likes
Increase in Community discussions	Suggested Actively being discussed in the Community (show to users who have historically participated in Community discussions)
Frequently deployed model	Suggested Popular choice in Text2Text Generation deployed to Inference Endpoints (show to Pro users or users who have previously deployed models to Inference Endpoints)

Highest affinity amongst creators you follow	You Like /TheBloke /TheBloke uploaded a model /MSFT uploaded a dataset /MetaAI created a Space (these are actions from creators I have the highest affinity with and should be ranked higher)
Out Of Network content reacted by high authority creators	Suggested /teknium uploaded a model (I don't follow /teknium but they have high authority and relevant to my interests)
Highest affinity amongst Topics you follow	Suggested Trending in LLMs (Activity such as a model, dataset etc. from a creator that is a top contributor to a Topic I follow)
Content from creators marked as favorite	From your favorites
Top content from creators in a specific location matches the users' location	Suggested Trending in New York (Show activity from creators popular amongst people from the same location as me)
Content from creators in "Follow your favorite AI Creators" carousel	Suggested Popular amongst people like you (Rank content created from creators in Recommended For You list and choose top k to insert into Home Feed with the appropriate justification text)
Content from trending items in popular Tasks	Suggested Trending in Text2Text Generation
Content being clicked on the most for a specific search term	Suggested Trending in Search

For You Feed

A For You feed is primarily focused on content and creator discovery. It may also be a solution for the cold-start problem.

While HF does have the Trending feed to show new users, it isn't personalized.

A FY feed for new users can be personalized by enhancing the onboarding process.

Step 1 - Personalization based on Onboarding

Current details required during sign-up and possible use of the information for personalization are;

- Username [N/A]
- Full name [N/A]
- Avatar (optional) [N/A]
- GitHub username (optional)
 - can lookup Github username and identify interests, repos starred to improve relevance.
 - No guarantee user is providing their own handle unless HF enables linking Github profile.
- Homepage (optional)
 - crawl the homepage to learn more about the user.
 - No guarantee of authenticity
- Twitter username (optional)
 - check Twitter bio, who they follow
 - Can link Twitter profile to find their community on HF
- AI & ML interests (optional)
 - Provide a collection of topics that a user can choose to reduce their effort of typing.
 - Allow users to choose Tasks that they are interested in. Provide a list of tags to choose from.

Suggestions

- LinkedIn profile
 - Let users provide their LinkedIn profile - use details from LinkedIn profile to personalize.
- **Choose Interests**
 - **Show users a carousel of popular creators.** These can be creators who have global appeal like /TheBloke or company profiles such as /Meta /MSFT
 - **Show users a carousel of popular models and datasets from each task type** such as LLama2 for text2text or SDXL for text2image
 - **Show users popular creators from their geographies** - e.g., /segmind is an Indian startup and has been capturing the imagination of the AI community in India. A signup from India would find relevance in seeing this suggestion during the onboarding flow.

Step 2 - construct feed based on choices made during onboarding

Details shared during onboarding (followed creators/companies/tasks) can be used immediately to construct a feed before the user lands on the home page.

News

A News feature on Hugging Face will provide industry context to users. Already, many organizations are on HF and these organizations also publish independent engineering blogs and post content on other social media sites like Twitter and LinkedIn.

Users may conserve effort by coming to HF for news related to the organizations they like.

Integrated Workflow for Data to Model

For many applications in NLU and Computer Vision a dataset is often created from scratch. This involves using tools like Label Studio for data annotations/labeling.

Once data is labeled, a model is trained and assessed. With LLMs, HITL is involved in post-training for evaluations using tools such as Agrilla.

Currently HF requires the integration of data labeling tools (e.g., Label Studio) and Notebooks for ad-hoc ML tasks (like Kaggle).

A research paper authoring solution would provide a complete workflow for researchers who can execute the Data -> Model -> Paper workflow on HF end-to-end.

Key Metrics

Metrics to track efficacy of social features fall into two groups.

Input Metrics

Input Metrics measure actions taken by the users on the product surface area.

Improved Creator Recommendations should lead to an increase in;

- Follows - I hit follow because the recommendation is relevant and the justification text tells me why I'm being recommended this creator/topic/task.
- Profile Visits - I am interested in learning more before I hit follow because the recommendation has piqued my interest due to its relevance.

Improvements in Feed relevance should lead to an increase in;

- Reactions - I like the content being shown to me.
- Increase in Feed scroll depth - I like the content and I keep scrolling for more.
- Profile visits - I click through to learn more about the creator/model/dataset/space.
- Follows - I follow creators because the suggested content in my feed is relevant.

Output Metrics

Due to an increase in relevance of the suggestions being made to the user, the user will exhibit the following changes in behavior that can be measured as output metrics.

- Time Spent on Product - the user will spend more time on the product and average time spent for cohorts will tend to increase.
- Retention - users will come back to the product and D1/D7/D30 retention numbers will go up. I reckon long term retention numbers are healthy for HF but not the short-term ones such as $\leq D7$.
- Participation Metrics may include users uploading models, creating spaces and such. Users converting from consumers to creators is an important metric to track.

This writeup was an attempt to apply my learnings from Koo to a vertical social network like Hugging Face.

About Harsh Singhal

Harsh developed ML-powered social media features on [Koo](#), the worlds' #1 multilingual microblogging platform (downloaded by 60M+ users in the last 3 years across 100 countries).

Harsh grew the ML team from 2 to 20 and shipped features such as Topics in 10+ languages, Recsys features such as Similar Creators, Recommended Creators, For You and developed Content Security features for platform safety.

Harsh has 16+ years of experience and has held Data Science and Machine Learning roles at LinkedIn, Netflix, Adobe and most recently at Glean.