Breakthrough Capital LLC

Identifying Emerging Technology Investment Targets

Table of contents

01 Background & Problem Statement

02 Proposed Solution

03 Methodology & Outcomes

04 Recommendations

Background & Problem Statement

01

Mission Statement: Stay ahead of the curve by identifying and investing in the most promising emerging technologies.

"snoo"

*Can we identify the the origin of Reddit posts **while** minimizing Type II errors so topics of interest (investment targets) are not missed?*

Reddit: A popular news and content **aggregation** site divided into communities (subreddits) by discussion topic.

Estimated ~55M daily users, projected to grow by 25% within 2 years

Target Subreddit

Subreddit 1: “Futurology”

A subreddit devoted to the field of Future(s) Studies and evidence-based speculation about the development of humanity, technology

- Titles and ideas aggregated may help identify areas for VC targets
- Popular links aggregated here may provide insight to other reliable news sources

https://www.bankmycell.com/blog/number-of-reddit-users/

https://unsplash.com/photos/man-wearing-black-top-2r5adxul49E

02

Proposed Solution

Strategic planning project plan

ID

Identify technology focused subreddit

Train

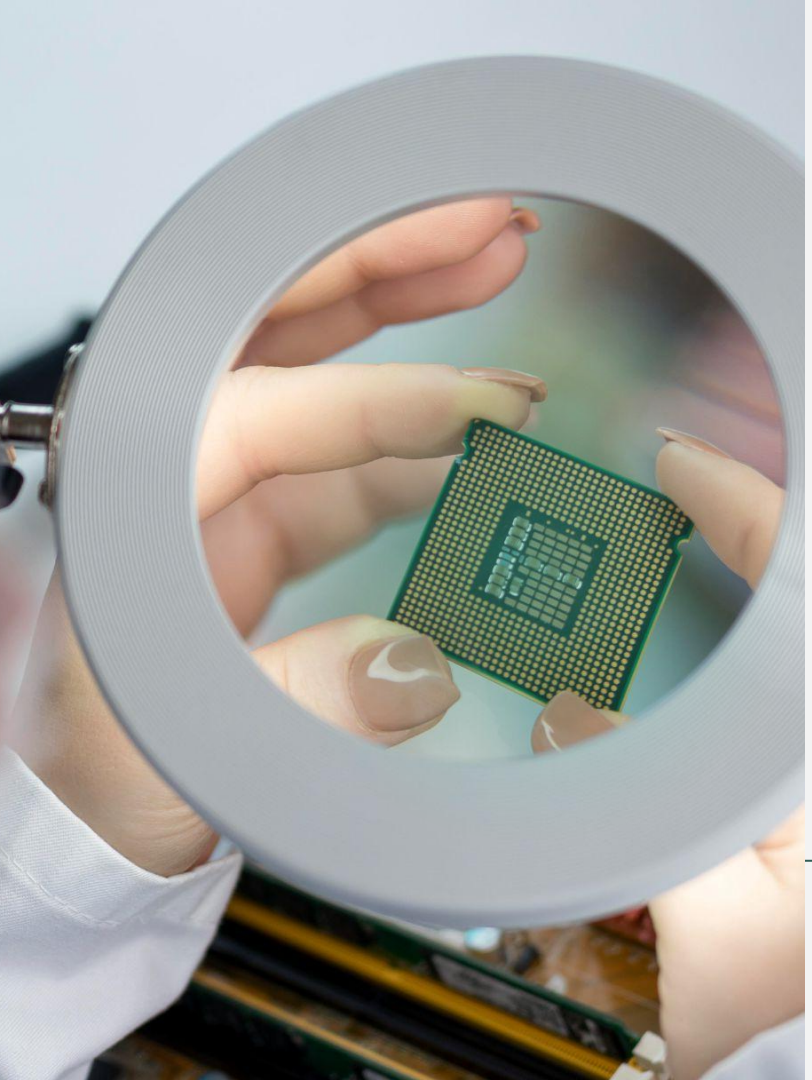
Train a model against subreddit with some overlapping themes

Optimize

Minimize false negatives (Type II) errors

Extract

Extract popular themes and domains



03

Methodology

Subreddit 1: “Futurology”

A subreddit devoted to the field of Future(s) Studies and evidence-based speculation about the development of humanity, technology

- 19M subscribers
- 1695 posts

Subreddit 2: “Scifi”

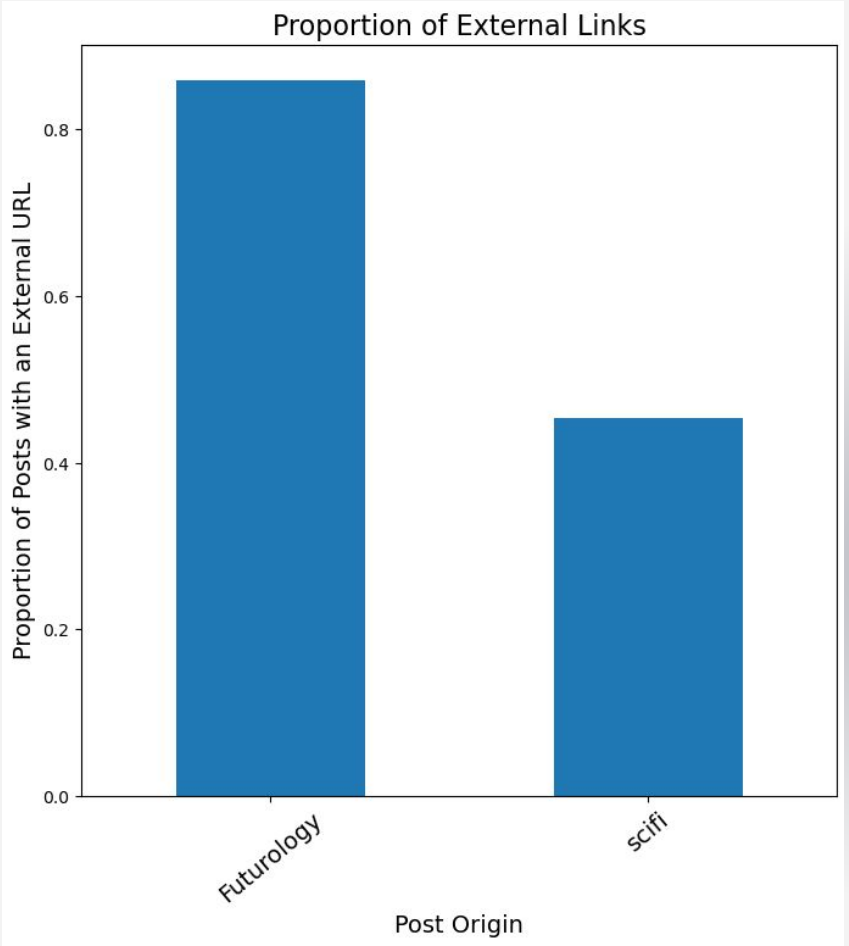
A subreddit centered around the Science Fiction or Speculative Fiction genre. “Laser guns, spaceships, and time travel”...

- 4.1M subscribers
- 1961 posts

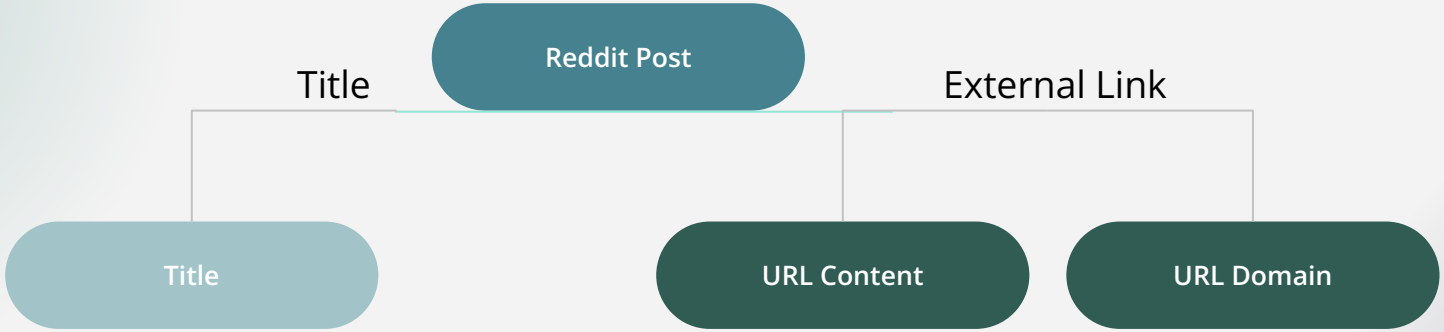
https://www.reddit.com/r/Futurology/
https://www.reddit.com/r/scifi/

Why URLs? Why no self text?

- Both subreddits feature a large amount of posts leading to external sites
- For this project, domains reddit.com and i.redd.it were ignored
- External links mean no self text, so titles were used instead



Feature Breakdown



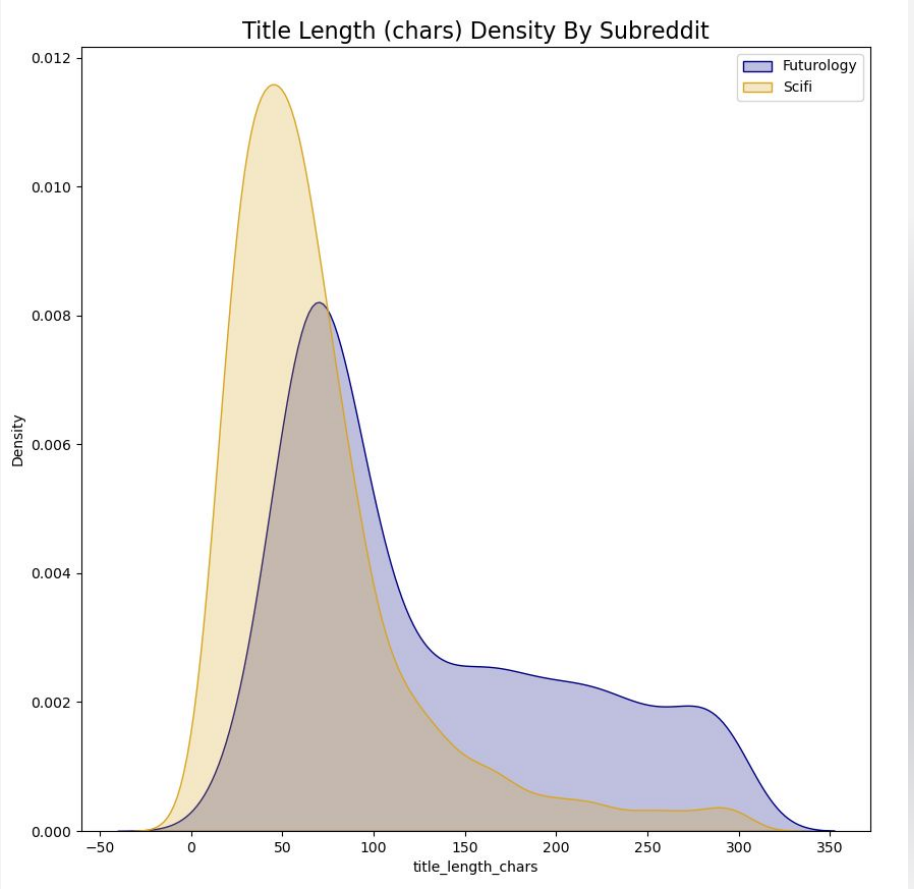
- Categorical
- Subject to NLP
  - WordNet Lemmatized
  - Vectorized

- Categorical
- Processed with custom URL tokenizer
- Subject to NLP
  - Vectorized

- Categorical
- Processed with domain extractor function
- OneHotEncoded

Interesting Aside - Why No Meta Features on Title like length?

- Initial impression was that they'd help based on visualizations like this one
- Removing it proved to increase both train and test score

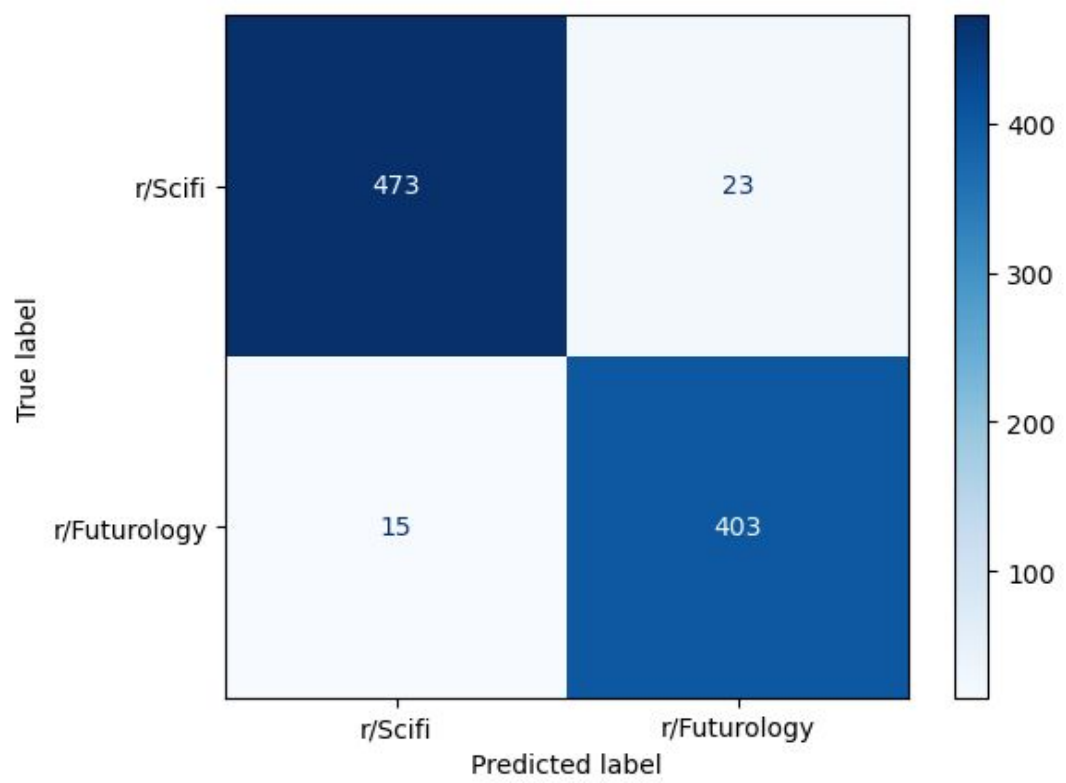


Model Analysis

Model	Transformers	Test Data Results
Multinomial Naive Bayes	Count Vectorize (Title, URL content) One Hot Encode (URL domain)	Accuracy: 95.8% **Recall: 96.4%** All Data Accuracy: 98% All Data Recall: 96.5%
Multinomial Naive Bayes	TFIDF Vectorize (Title, URL content) One Hot Encode (URL domain)	**Accuracy: 95.9%** Recall: 95.9%
Random Forest	Count Vectorize (Title, URL content) One Hot Encode (URL domain)	Accuracy: 91.5% Recall: 88%
C Support Source Vector Classification	Count Vectorize (Title, URL content) One Hot Encode (URL domain)	Accuracy: 92.5% Recall: 88.9%

Record for most expensive: Model 2 - with 155,521 models elapsing 37m,
"Friends don't let friends GridSearch"

Confusion Matrix - Holdout Data

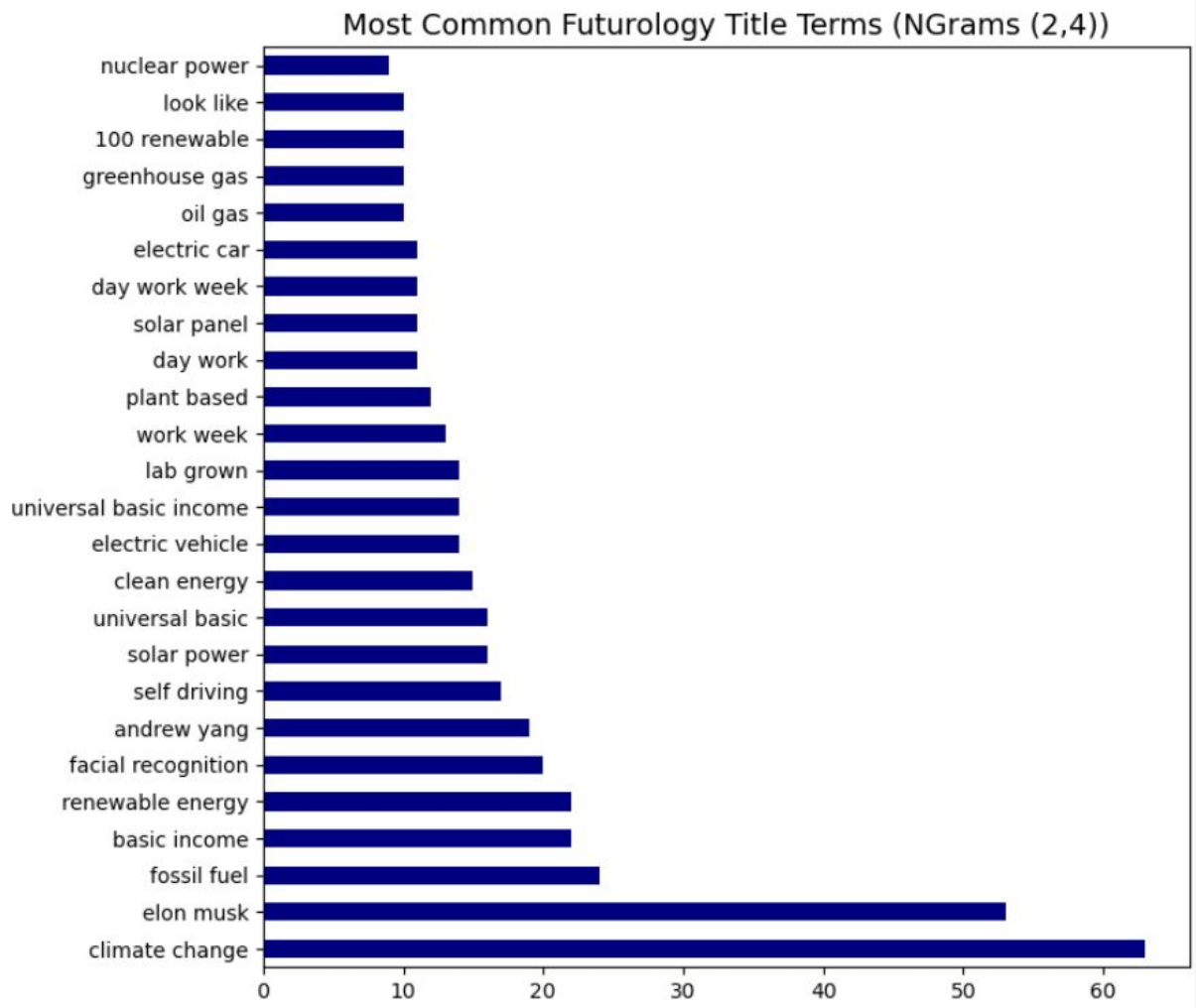


Optimized Type II Errors

- Model had highest recall; least likely to misclassify a true Futurology post
- Investigating false positives is cheap, missing potential leads is not

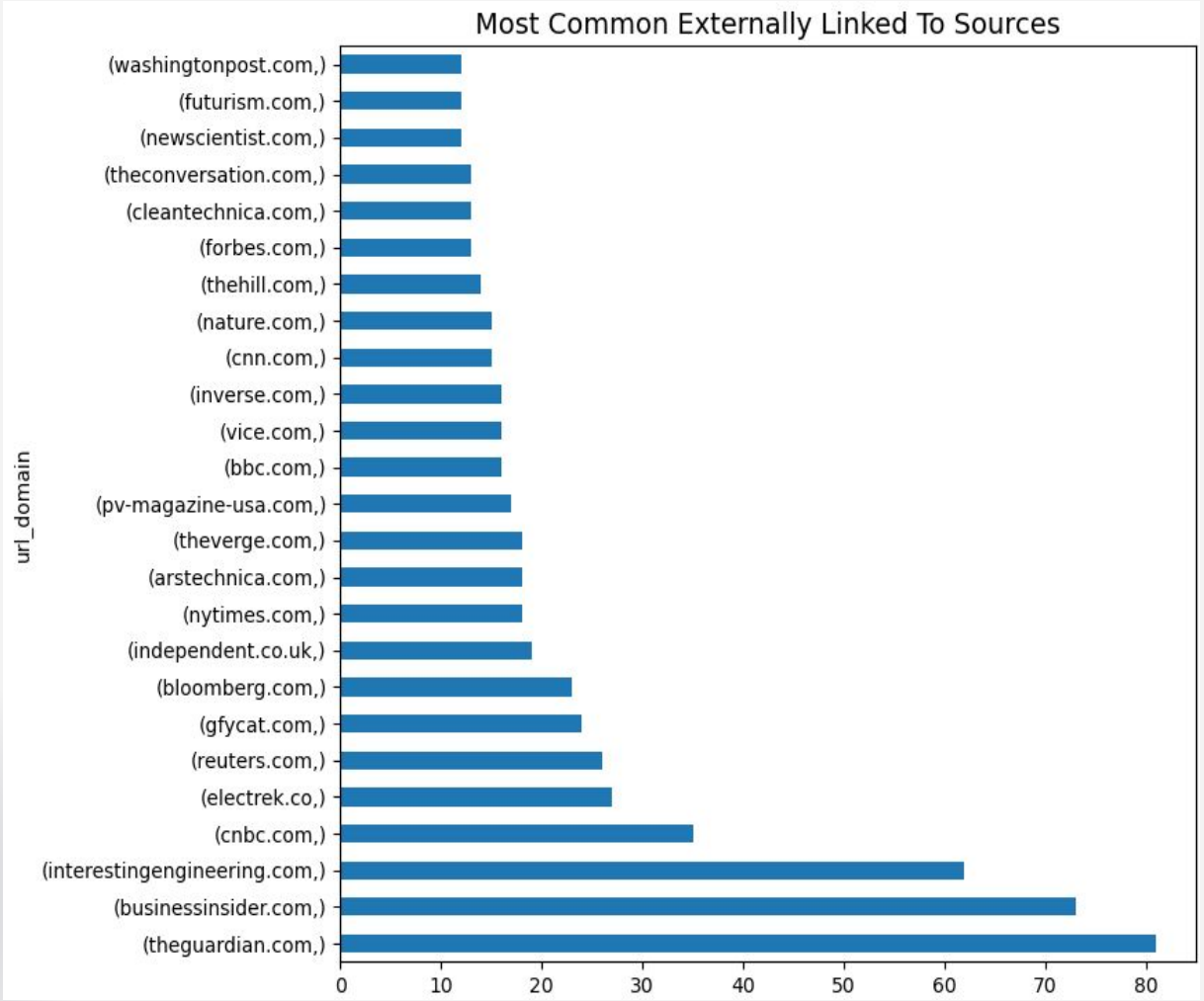
04

Insights and Recommendations



Invest in the future:

- Increase research into market growth and startup activity focused on areas like:
  - Climate change
  - Renewable energy
  - Facial recognition
  - Universal basic income effects
  - Self driving cars
  - Lab grown protein
  - etc



Stay on the pulse:

- Increase monitoring of these websites for futurism articles gaining traction:
  - The Guardian
  - Business Insider
  - Interesting Engineering
  - Cnbc Electrik
  - Reuters
  - Bloomberg
  - etc.

For your amusement - best of the false positives (Predicted Futurology but were Scifi)

What technologies could potentially be invented in the near future?

For your amusement - best of the false negatives (Predicted Futurology but were Scifi)

What technologies could potentially be invented in the near future?

Scientists at Harvard and MIT have collaborated to discover an entirely new form of matter. They have created molecules from light - and the properties of these new photon-molecules are similar to that of a LIGHT SABER

For your amusement - best of the false negatives (Predicted Futurology but were Scifi)

What technologies could potentially be invented in the near future?

Scientists at Harvard and MIT have collaborated to discover an entirely new form of matter. They have created molecules from light - and the properties of these new photon-molecules are similar to that of a LIGHT SABER

According to Nikola Tesla, Pyramids in the Egypt were giant transmitters of energy. What's your opinion? Is it true?

For your amusement - best of the false negatives (Predicted Futurology but were Scifi)

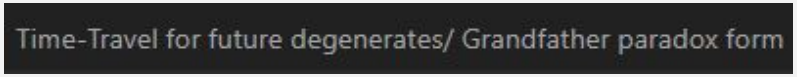
What technologies could potentially be invented in the near future?

Scientists at Harvard and MIT have collaborated to discover an entirely new form of matter. They have created molecules from light - and the properties of these new photon-molecules are similar to that of a LIGHT SABER

According to Nikola Tesla, Pyramids in the Egypt were giant transmitters of energy. What's your opinion? Is it true?

How to cause a global dark age?

For your amusement - best of the false negatives (Predicted Scifi but were Futurology)



For your amusement - best of the false negatives (Predicted Scifi but were Futurology)

Time-Travel for future degenerates/ Grandfather paradox form

Why are there so many anomalies in the universe?

For your amusement - best of the false negatives (Predicted Scifi but were Futurology)

Time-Travel for future degenerates/ Grandfather paradox form

Why are there so many anomalies in the universe?

What interplanetary wars might look like in the future

For your amusement - best of the false negatives (Predicted Scifi but were Futurology)

Time-Travel for future degenerates/ Grandfather paradox form

Why are there so many anomalies in the universe?

What interplanetary wars might look like in the future

What will music sound like in 2024

For your amusement - best of the false negatives (Predicted Scifi but were Futurology)

Time-Travel for future degenerates/ Grandfather paradox form

Why are there so many anomalies in the universe?

What interplanetary wars might look like in the future

What will music sound like in 2024

Why not everyone on the world not trying to make a sci-fi utopia?

For your amusement - best of the false negatives (Predicted Scifi but were Futurology)

Time-Travel for future degenerates/ Grandfather paradox form

Why are there so many anomalies in the universe?

What interplanetary wars might look like in the future

What will music sound like in 2024

Why not everyone on the world not trying to make a sci-fi utopia?

Question about cryogenically freezing people

For your amusement - best of the false negatives (Predicted Scifi but were Futurology)

Time-Travel for future degenerates/ Grandfather paradox form

Why are there so many anomalies in the universe?

What interplanetary wars might look like in the future

What will music sound like in 2024

Why not everyone on the world not trying to make a sci-fi utopia?

Question about cryogenically freezing people

It's not aliens. It'll probably never be aliens. So stop. Please just stop.

Presentation possible via