

アドバンス・シミュレーション 第6回 ・セミナー 2023

2023年 8 月28日(月) 離

プログラム

1. アドバンスソフト株式会社についてと、先生のご紹介 1

招待講演

2. **「フェイクニュースと生成 AI」** 3
東京工業大学 環境・社会理工学院 准教授 笹原 和俊 様

3. アドバンスソフトからの情報提供 29

講演概要

「フェイクニュースと生成 AI」

東京工業大学 環境・社会理工学院 准教授 笹原 和俊 様

この講演では、生成 AI によってさらに進化することが予想されるフェイクニュースやディープフェイクの問題について焦点を当てる。ChatGPT や Stable Diffusion などの生成 AI の急速な技術的發展、情報の真偽を判断するための戦略や、デジタルプラットフォーム上で情報の拡散を防ぐ技術的対策についても説明する。その上で、これら複雑な問題に対する対策を立てるためにはビッグデータの活用が必要不可欠であり、それによって得られた結果をもとに、社会シミュレーションを用いて具体的な対策を模索することの重要性についても触れる。

最後に、デジタル社会における生成 AI の可能性と課題について議論する。



アドバンスソフト株式会社 セミナー事務局

〒101-0062 東京都千代田区神田駿河台四丁目3番地 新お茶の水ビルディング 17階西

TEL: 03-6826-3971 FAX: 03-5283-6580

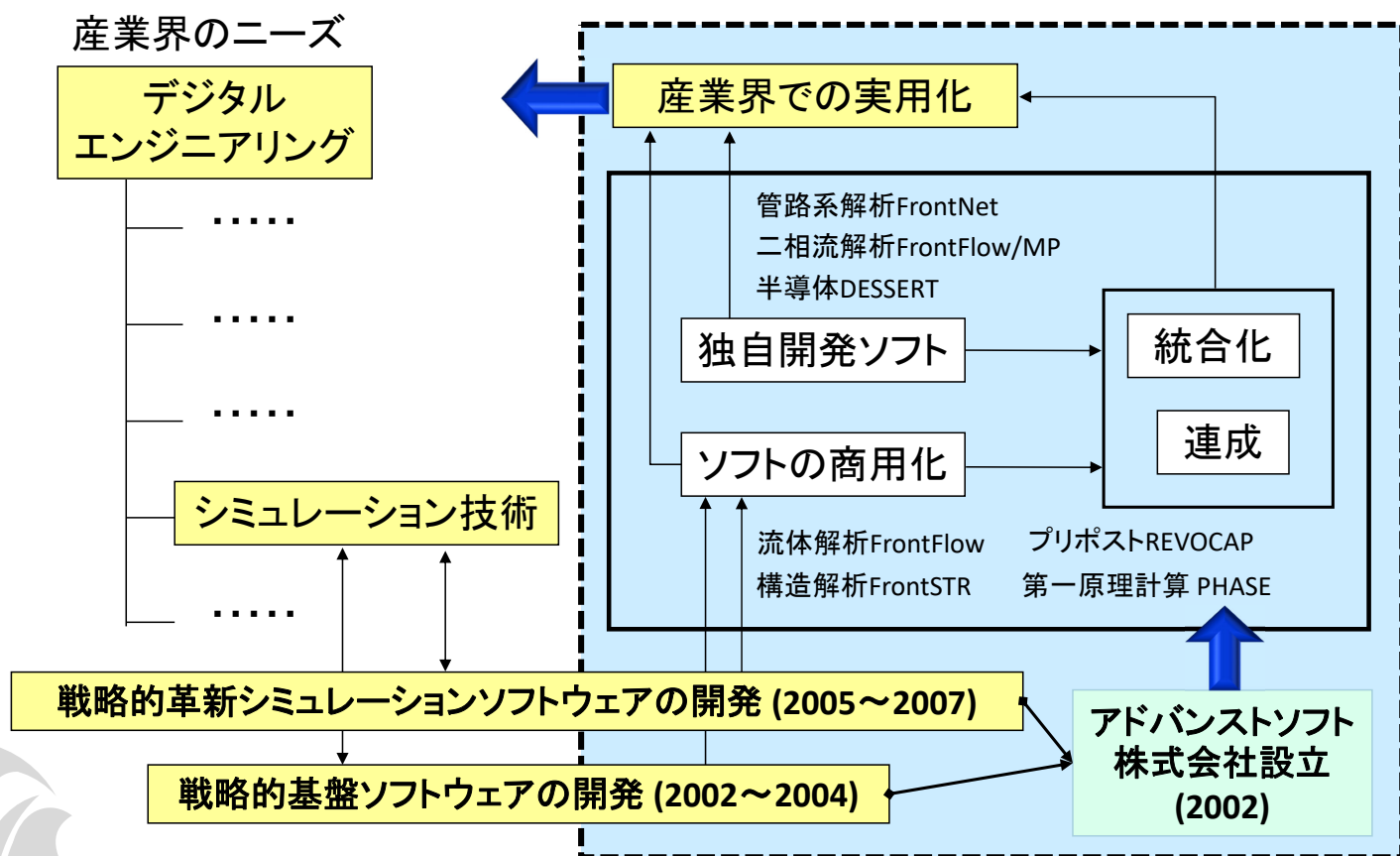
URL: <http://www.advancesoft.jp/> E-mail: office@advancesoft.jp

アドバンスソフト株式会社についてと 先生のご紹介

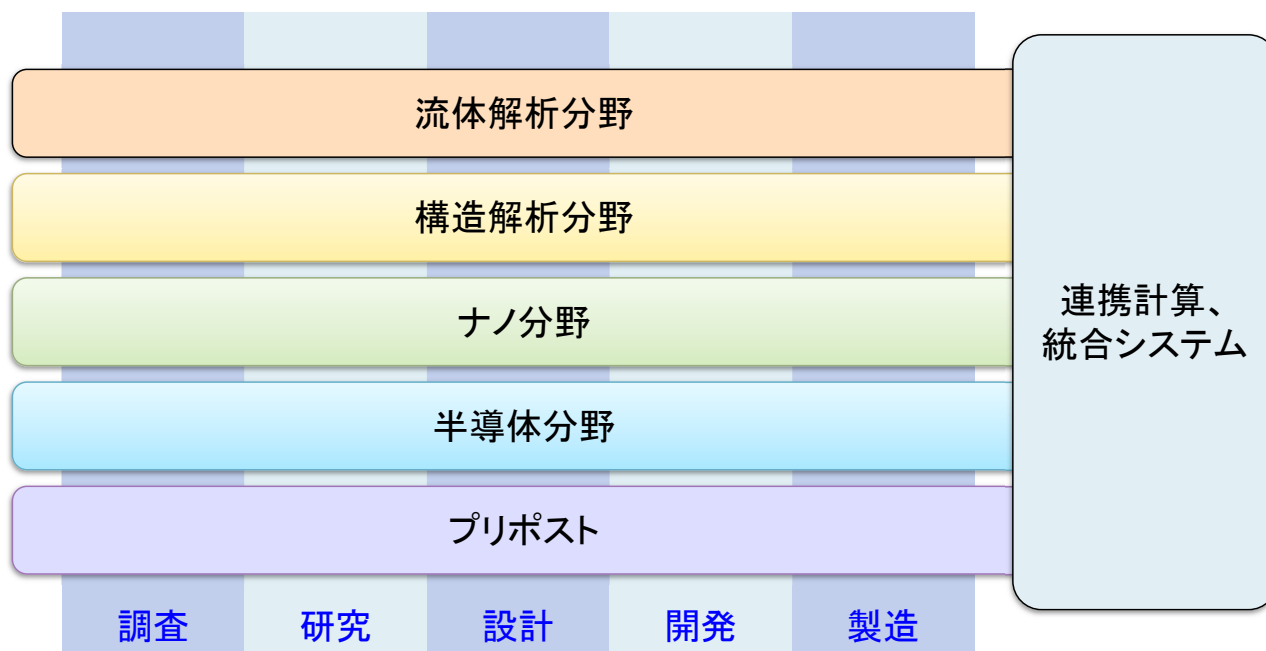
第6回 アドバンス・シミュレーション・セミナー
(東京工業大学 環境・社会理工学院
准教授 笹原 和俊 様)

2023年8月28日 (月) 開催
アドバンスソフト株式会社

アドバンスソフトとは



事業分野



産業の主要な分野のあらゆるフェーズで直面する課題に対し、
科学技術計算によるソリューションをご提供します。

笹原 和俊 先生 のご紹介

ご経歴

2005年 東京大学大学院 総合文化研究科修了。博士（学術）。
名古屋大学大学院 情報学研究科助教・講師を経て、現在、東京工業大学環境・社会理工学院
准教授。国立情報学研究所 客員准教授。

ご研究内容

専門は、計算社会科学。主著に『フェイクニュースを科学する』（化学同人）、『ディープ
フェイクの衝撃』（PHP研究所）がある。

第6回 アドバンス・シミュレーション・セミナー (2023/8/28)

フェイクニュースと生成AI

笹原 和俊

東京工業大学 環境・社会理工学院

自己紹介

- 1976 福島県生まれ
- 2005 東京大学 大学院総合文化研究科修了 (博士 (学術))
- 2012～2020 名古屋大学 大学院情報学研究科 助教・講師
- 2016～2020 JSTさきがけ研究者 (兼任)
- 現在 東京工業大学 環境・社会理工学院 准教授
国立情報学研究所 客員准教授
- 専門 計算社会科学

発表内容

- フェイクニュースとはどのような問題か
- エコーチェンバーとフィルターバブル
- ディープフェイクと生成AI
- 高度化するフェイクに対抗する

3

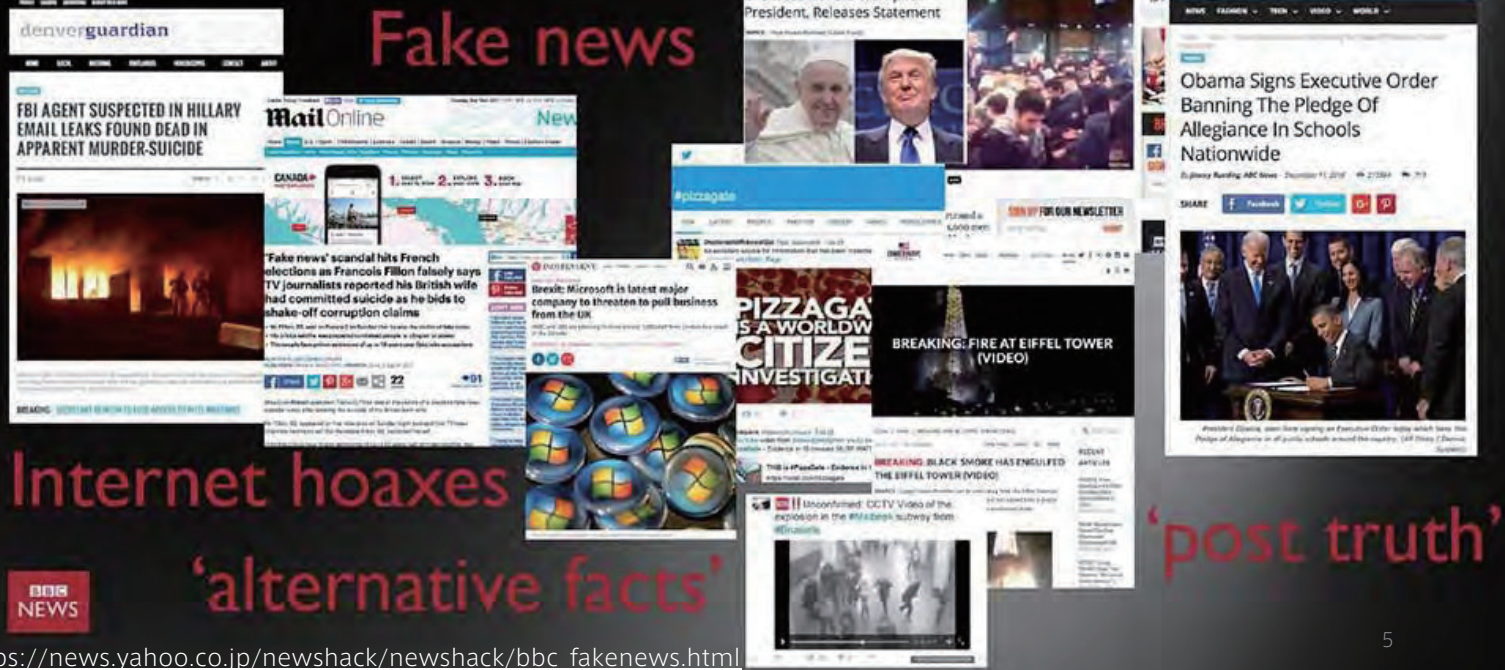
発表内容

- フェイクニュースとはどのような問題か
- エコーチェンバーとフィルターバブル
- ディープフェイクと生成AI
- 高度化するフェイクに対抗する

4

フェイクニュースの氾濫

We are living in uncertain, confusing times – when it can be hard to know what to believe



5

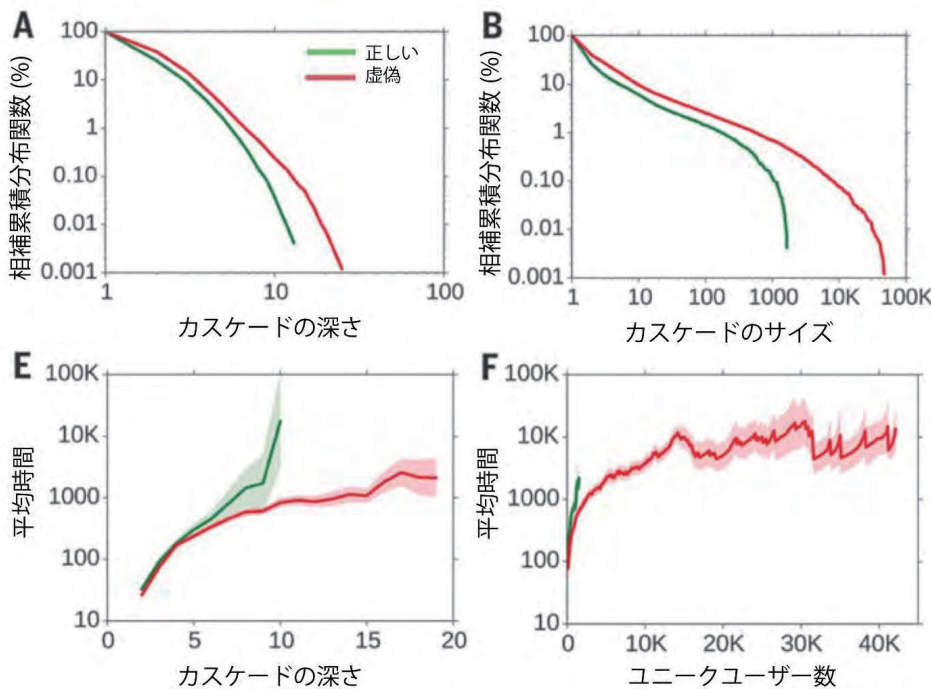
フェイクニュースの10類型

プロパガンダ <ul style="list-style-type: none"> 政府、企業、NPOなどが、人の意識や価値感、知識に影響を与えるための手段 感情に訴えてくる 利益になることもあれば、害を及ぼすこともある 	党派的情報 <ul style="list-style-type: none"> イデオロギー的で事実の解釈を含むが、中立を装っている 自分たちに都合の良い事実を強調し、それ以外は取り上げない 感情的で情熱的な言葉を使う 	影響度 <ul style="list-style-type: none"> 場合による 低 中 高
釣りタイトル <ul style="list-style-type: none"> 本質から外れた、刺激的で目立つ見出し 見出しが内容を反映しておらず、誤解を与えやすい 広告収入を得るために利用される 	陰謀論 <ul style="list-style-type: none"> 恐怖や不確実さから、複雑な現実を単純化して説明しようとする 間違いだと証明しづらく、反証すると陰謀論の信頼性が増してしまう 専門家や当局情報を否定する 	
スポンサードコンテンツ <ul style="list-style-type: none"> 記事に見せかけた広告 報道機関と利害の対立を生むことがある 明示されていないければ、広告だと見抜くのは難しい 	ニセ科学 <ul style="list-style-type: none"> 見せかけの環境保護活動、奇跡の治療法、ワクチンの拒否、地球温暖化の否定 正しい科学的研究を、大げさな、またはうそその情報でねじ曲げる たいてい専門家の意見と矛盾する 	動機 <ul style="list-style-type: none"> 金銭 政治/権力 ユーモア/冗談 情熱 (誤)情報の伝達
風刺、架空の話 <ul style="list-style-type: none"> 社会批判またはユーモア 内容はさまざまで、意図が明確でないことも 事実と混同され、読者を困惑させる 	誤情報 <ul style="list-style-type: none"> 事実と間違いが入り交じったコンテンツ 情報を伝えたいと思っているが、作成者が誤りに気づいていないこともある 誤った引用、不適切に加工されたコンテンツ、誤解を招く見出し 	
誤報 <ul style="list-style-type: none"> 定評ある報道機関も間違えることがある 誤報はブランドを傷つけ、怒りを買う訴訟になることも 信頼できる報道機関であれば、誤りを認めてお詫びする 	偽情報 <ul style="list-style-type: none"> 人をだます目的で広く拡散する、完全なねつ造コンテンツ ゲリラマーケティング戦略、ポット、コメント、なりすましブランド 広告収入目当てか、政治的な影響を与える目的、または両方 	

<https://eavi.eu/beyond-fake-news-10-types-misleading-info/>

6

偽ニュースは速く遠くまでたくさん伝わる



拡散されやすい話題

- 政治
- 都市伝説
- ビジネス
- テロ・戦争
- 科学
- エンタメ
- 自然災害

Vosoughi et al. Science, 2018

7

新型コロナとインフォデミック



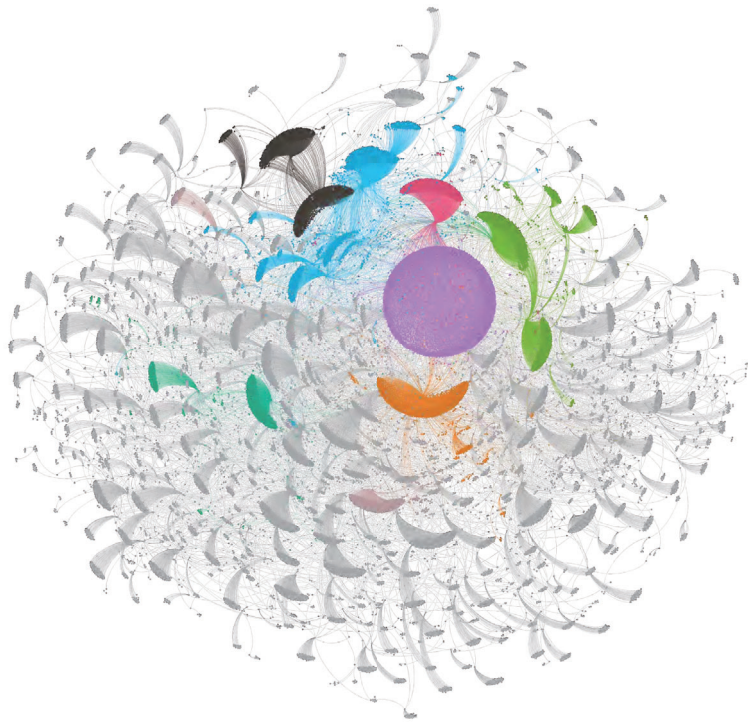
確かな情報と不確かな情報が混在する情報過多

- 間違った意思決定
- 不安や恐怖が増幅
- 差別を助長

出典：1st WHO Infodemiology Conference (Jun 30 - Jul 16 2020)

8

5G陰謀論の拡散

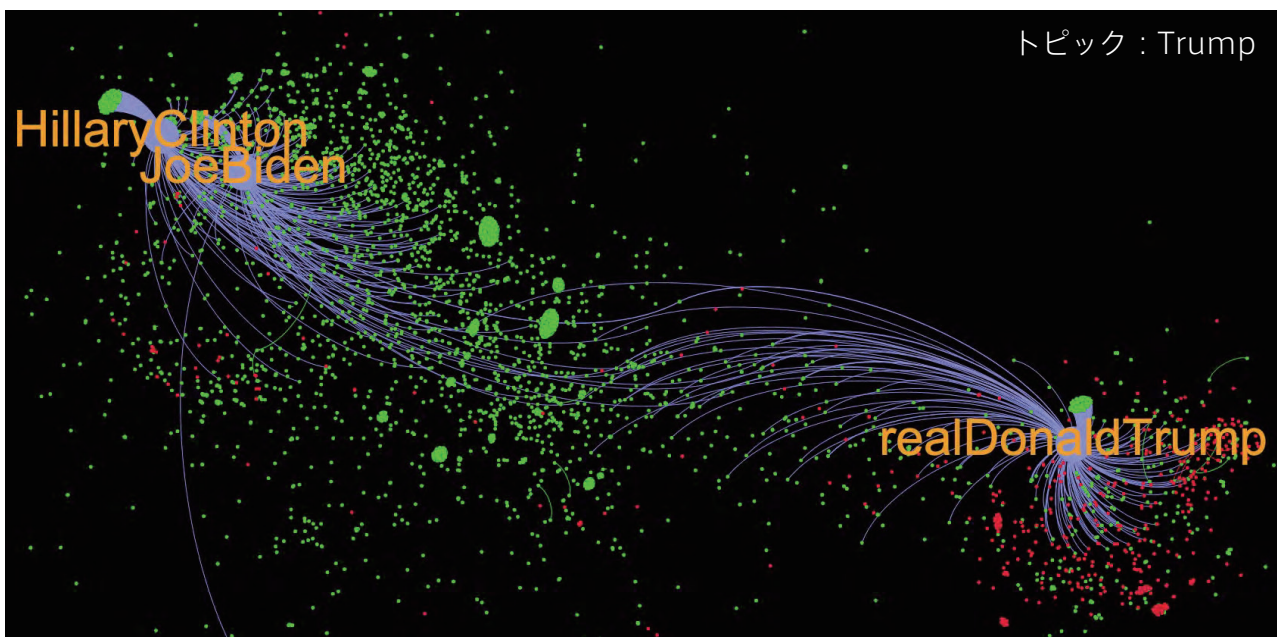


笹原和俊「デマや陰謀論はなぜネット上を拡散するのか」現代化学 (2020) 9

陰謀論を増幅するBot

赤：悪質なBot
 緑：普通のBot

トピック：Trump

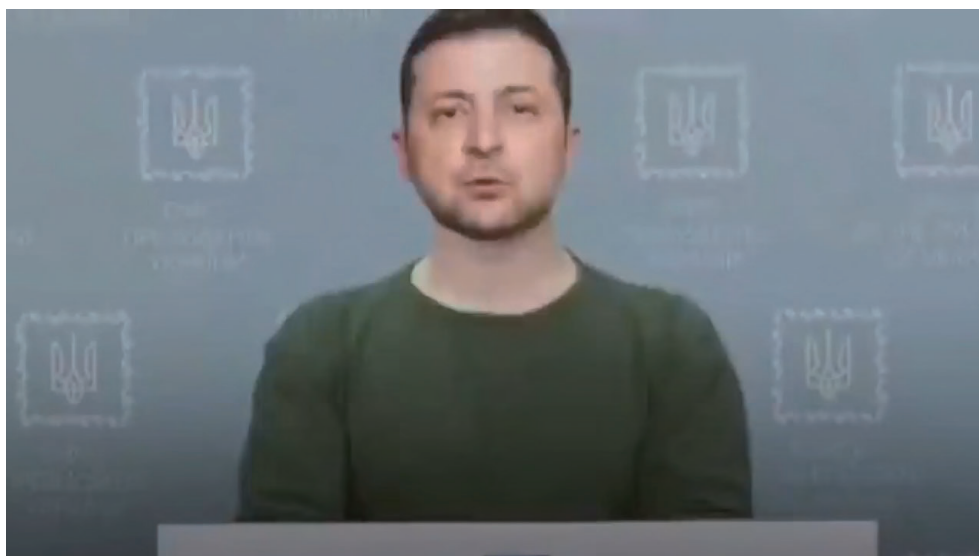


W. Xu and K. Sasahara, Journal of Computational Social Science (2022)

米大統領経済報告2023で引用

ディープ・ゼレンスキー

瞬きに注目！



<https://www.youtube.com/watch?v=X17yrEV5sl4>

13

発表内容

- フェイクニュースとはどのような問題か
- エコーチェンバーとフィルターバブル
- ディープフェイクと生成AI
- 高度化するフェイクに対抗する

14

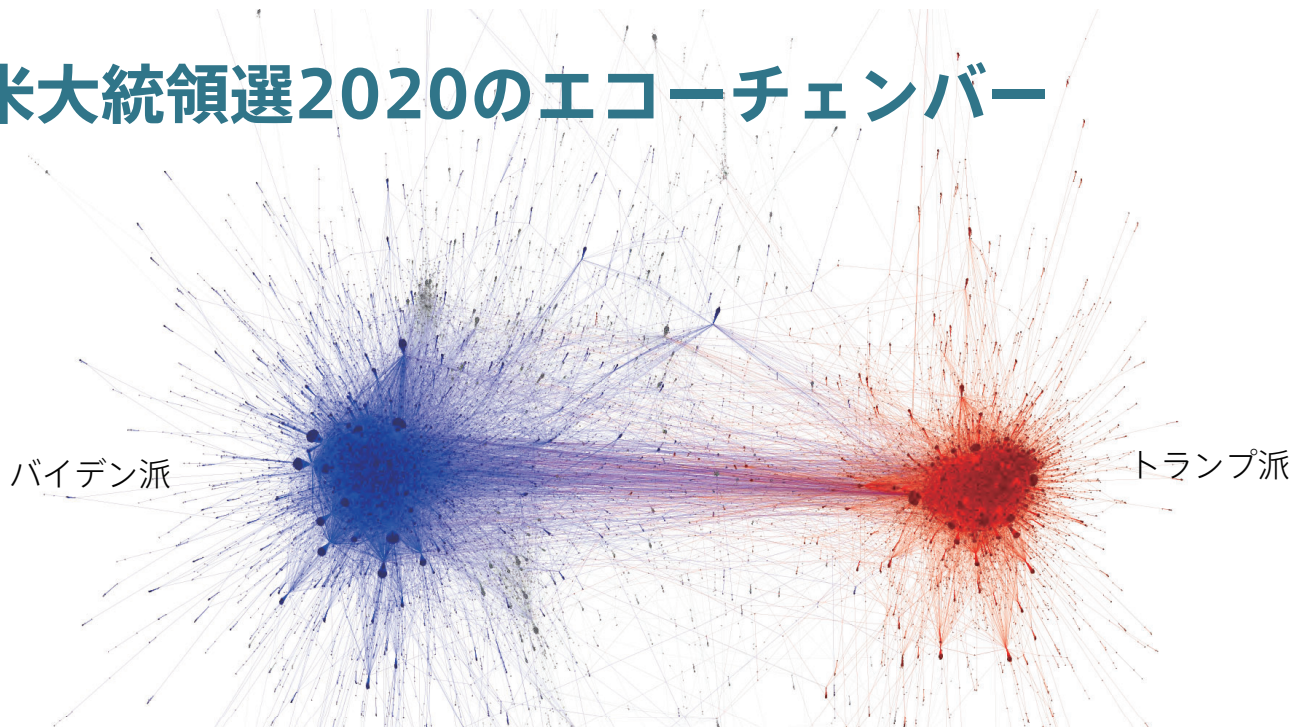
エコーチェンバーとは何か

同じ意見をもつ人々が集まり、自分たちの意見を強化し合うことで、多様な視点に触れることができなくなる現象

- 音が反響する物理現象のメタファー
- インターネットの文脈ではCass Sunsteinが提起 (2001)
 - 意見の二極化と社会的分断 (懸念)
- ホモフィリー (同類性) による選択的接触

15

米大統領選2020のエコーチェンバー



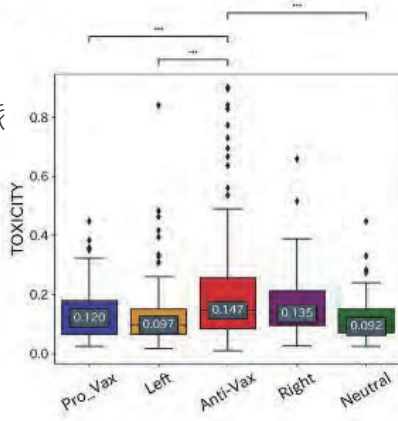
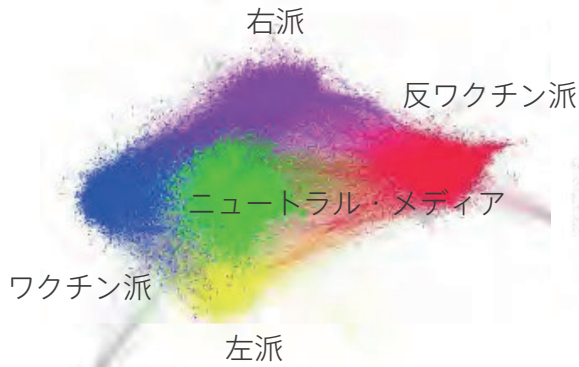
ツイートの拡散に見るリベラル系 (青) と保守系 (赤) のイデオロギーの分断

日本経済新聞電子版 (2020/10/30, 笹原作成)

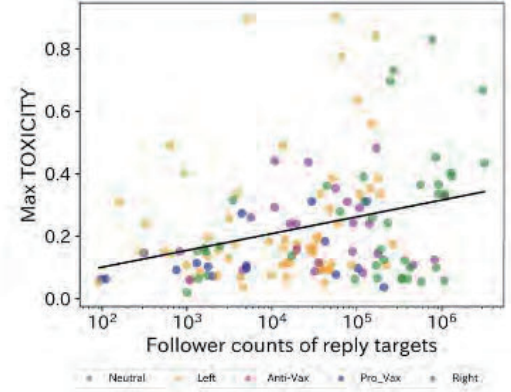
17

コロナ禍の反ワクチン運動

Japanese tweets



フォロワー数が多いほど毒性が高い
リプライを受け取る



K. Miyazaki et al. Humanities and Social Sciences Communications (2022)



NHK 「8Kがいざなう実世界展」 @渋谷スクランブルスクエア14F (2022/1/29~2/25)

SNSを模した意見形成モデル

- ソーシャルネットワーク $G = (V, E)$
 - V : エージェント、
 - E : 社会的つながり (フォロー)
- 意見 $o_i(0) \in [-1, 1]$ と許容範囲 ε
- 近しい意見 $\{m_j \text{ s.t. } |o_i - m_j| < \varepsilon\}$
- ソーシャルフィード：
 - 友人が(再)投稿した直近 l 個のメッセージを表示

22

SNSを模した意見形成モデル (続き)

社会的影響 (μ)

$$o_i(t+1) = o_i(t) + \mu \frac{\sum_{j=1}^l I_\varepsilon(o_i(t), m_j(t))(m_j(t) - o_i(t))}{\sum_{j=1}^l I_\varepsilon(o_i(t), m_j(t))}$$

社会的つながりの改変 (q)

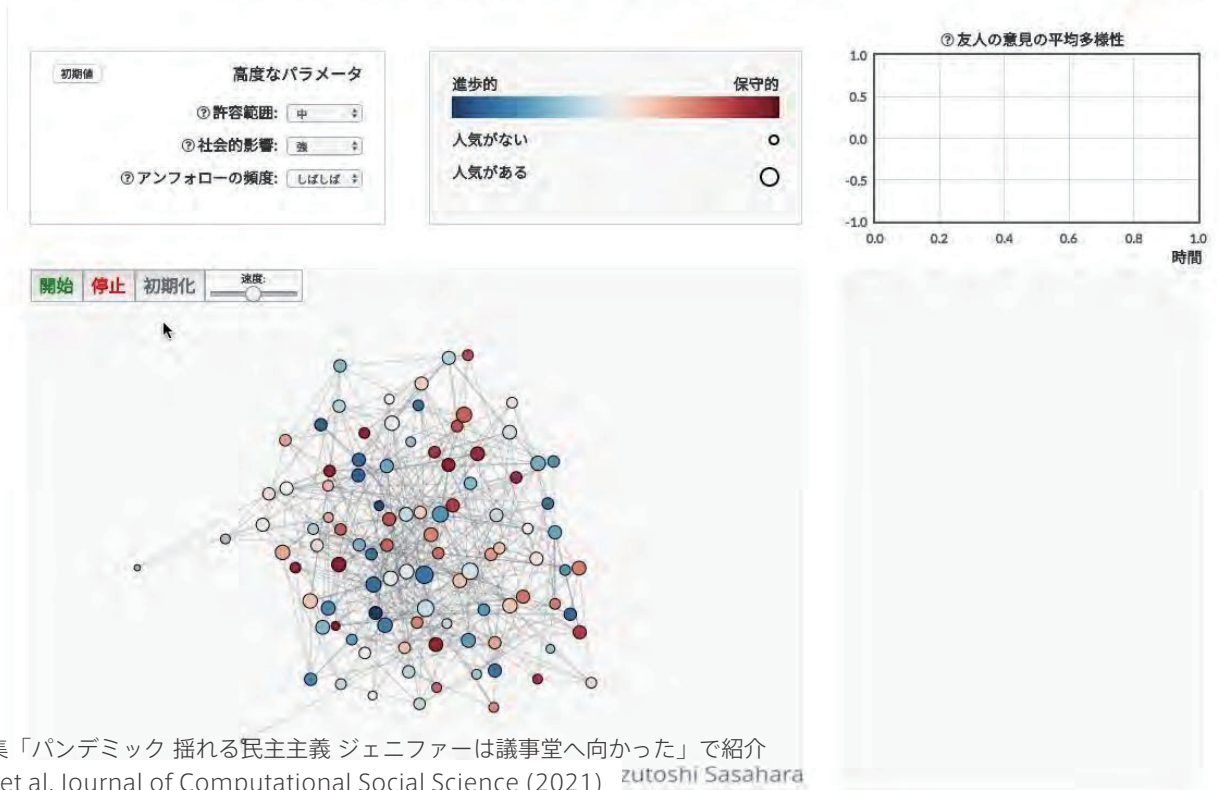
- 近しくないメッセージを投稿した友人をアンフォローする
- 新しい友人をいずれかの方法でフォローする：
 - 誰でもいい or ソーシャルフィードから or 友人推薦に従う

メッセージを投稿

- 自分の意見を投稿 ($1 - p$)
- ソーシャルフィード上の近しい意見を再投稿 (p)

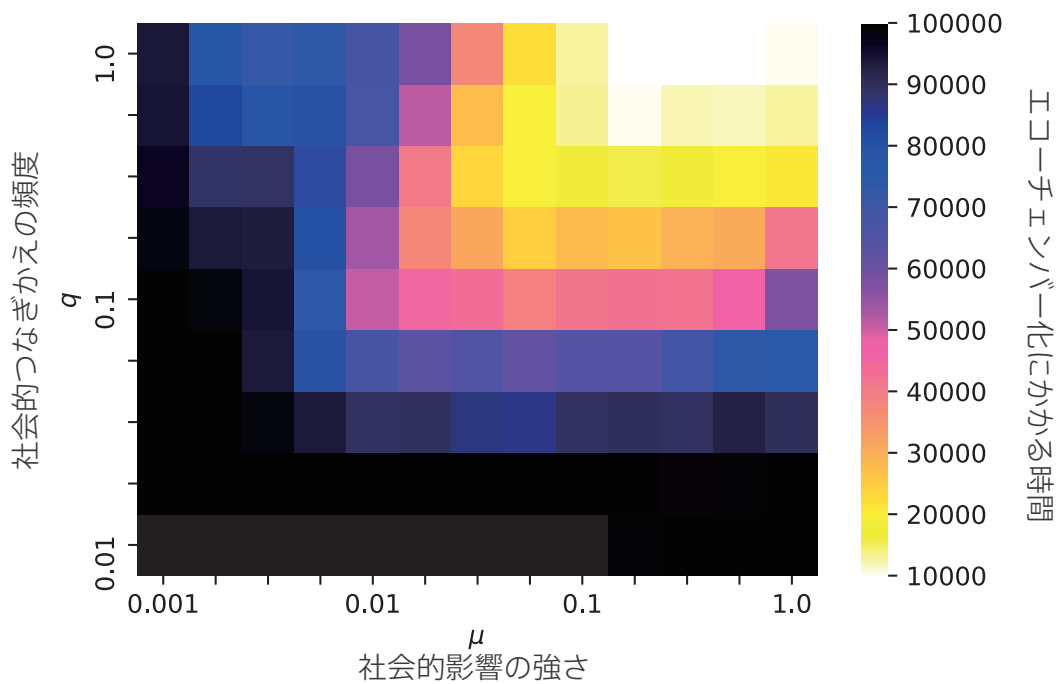
23

エコチェンバーはどのようにしてソーシャルメディアから生じるのか



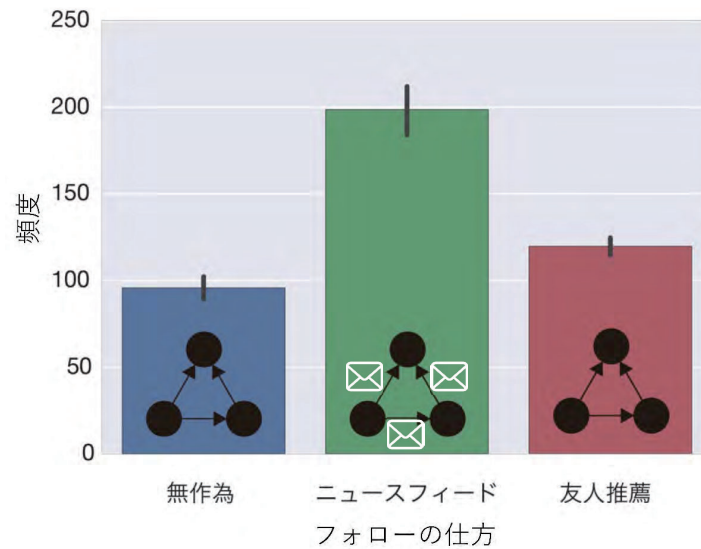
NHK ETV特集「パンデミック 揺れる民主主義 ジェニファーは議事堂へ向かった」で紹介
K. Sasahara et al. Journal of Computational Social Science (2021) zutoshi Sasahara

SNSはエコチェンバーを加速する



K. Sasahara et al. Journal of Computational Social Science (2020)

三者閉包=こだま化装置の増加



SNSのフォローの仕方は三者閉包を増加させ、それによって意見がこだま化

K. Sasahara et al. Journal of Computational Social Science (2020)

27

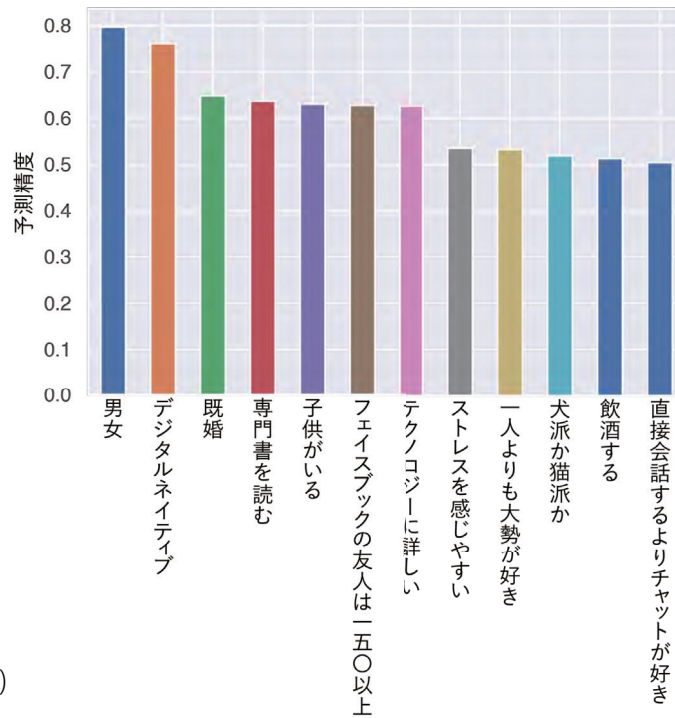
フィルターバブルとは何か

インターネット上の情報が、個々のユーザの関心や過去の検索履歴、閲覧履歴などに基づいて選択的に表示される現象

- 自分の価値観に合った情報にばかり触れることで、異なる視点に触れる機会が減少し、偏った認識や固定観念を持つ
- 「フィルター」のメタファー
- Eli Pariserが提起 (2011)
 - 知的孤立と社会的分断の懸念
- アルゴリズムによる選択的接触

30

ツイートが運ぶ個人属性

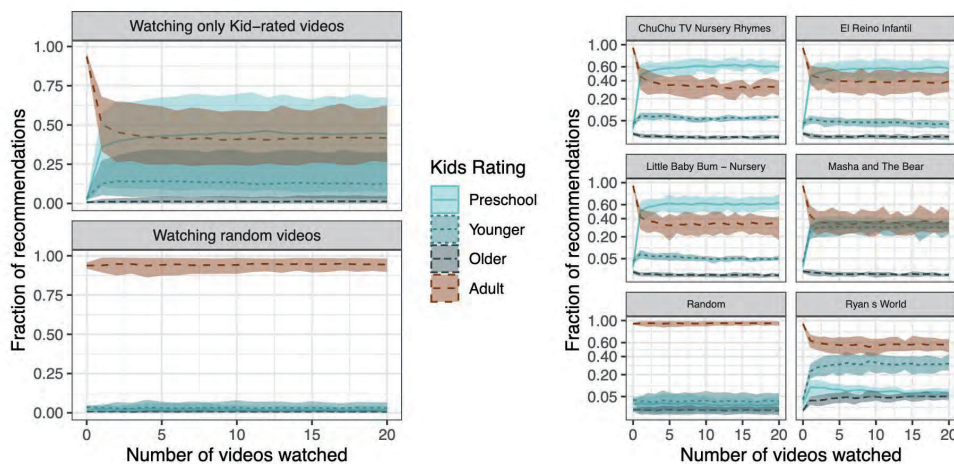


Yo and Sasahara (in prep.)

32

おすすめでラビットホールに落ちる

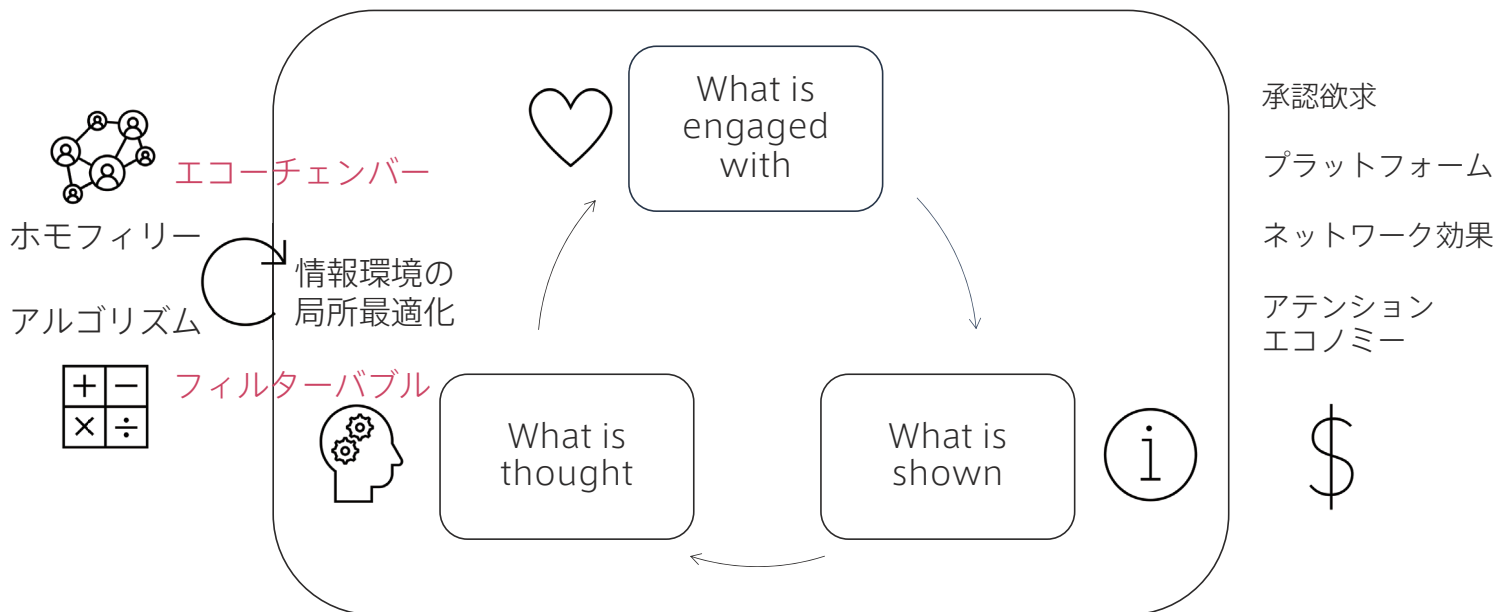
- 推薦された子供向け動画を2,3回見ると、そればかりが推薦されるように (ランダム視聴ではそのような傾向はなし)



Le Merrer, E. & Trédan, G., hal-03620039 (2022)

34

情報接触が制約される環境



36

発表内容

- フェイクニュースとはどのような問題か
- エコーチェンバーとフィルターバブル
- ディープフェイクと生成AI
- 高度化するフェイクに対抗する

37

どちらが本物でしょう？



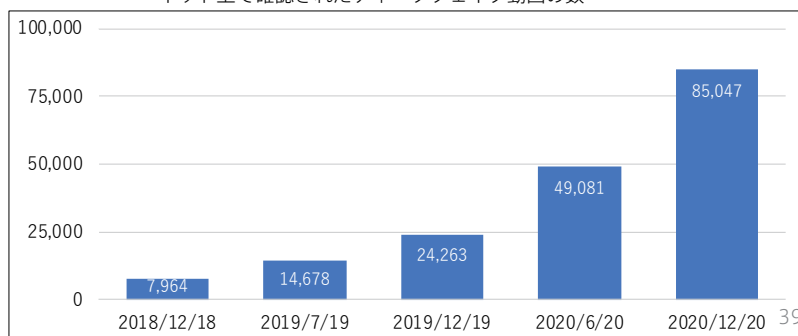
<https://www.whichfaceisreal.com/>

38

ディープフェイクとは

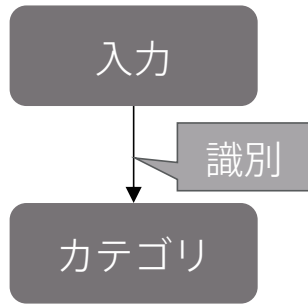
- ディープラーニング（深層学習）とフェイク（偽物）を組み合わせた造語
- 広義：AIによって合成・生成されたメディアやその技術
(cf. シンセティックメディア)
- 狭義：人をだます目的で、写真、音声、映像の一部を入れ替えて本物そっくりに合成された偽メディア

ネット上で確認されたディープフェイク動画の数

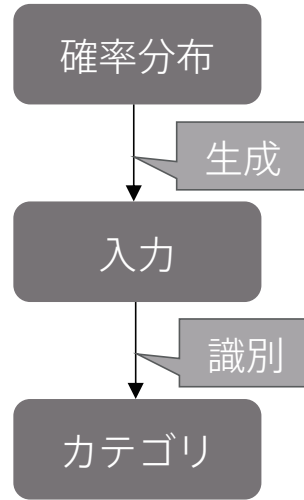


識別モデルと生成モデル

入力データが与えられたときに、それをカテゴリに分類する



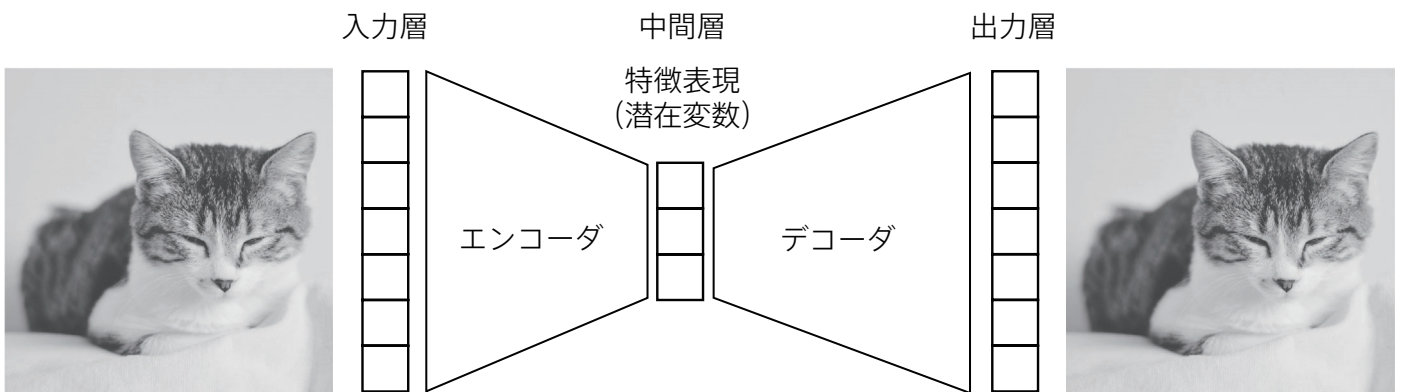
識別モデル



生成モデル

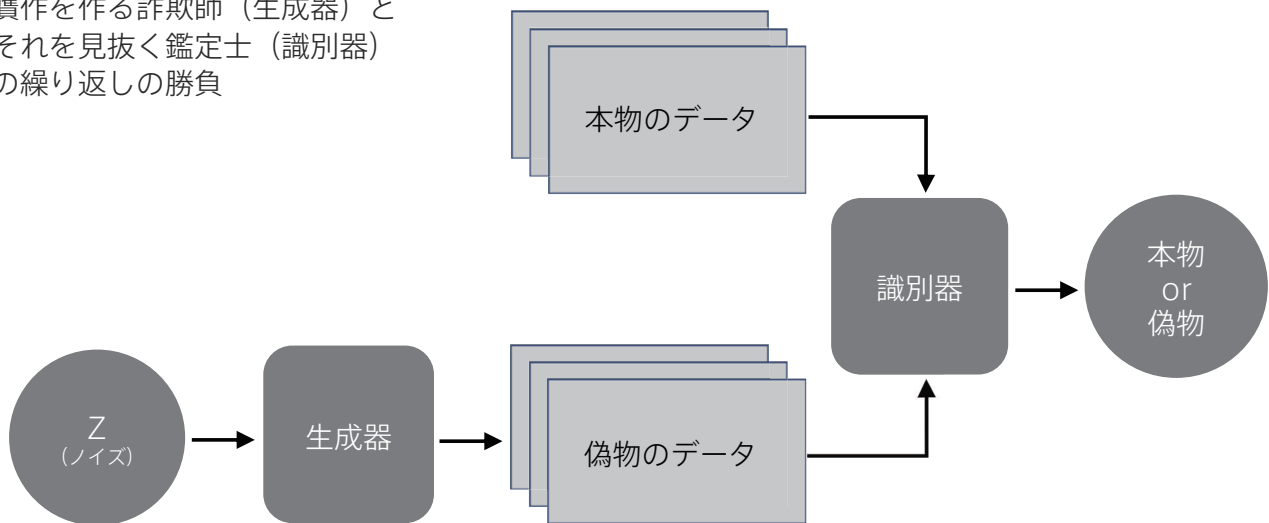
訓練データの特徴を捉えた新しいデータを生成する

特徴を学習するオートエンコーダ



GAN (Generative Adversarial Networks; 敵対的生成ネットワーク)

贋作を作る詐欺師（生成器）と
それを見抜く鑑定士（識別器）
の繰り返しの勝負



42

ウマをシマウマに

CycleGANによる画像スタイル変換



43

画像生成AI

- 言葉による指示（プロンプト）から画像を生成
 - これまではコンピュータの知識が必要だった
- DALL・E2 (Open AI)
- Midjourney
- Stable Diffusion (Stability AI)

FBIが1月6日連邦議会議事堂襲撃を組織
(FBI organized January 6 Capitol Riot)



画像生成AIによるフェイクニュース

ドローンで撮影された静岡県の水害。
マジで悲惨すぎる...

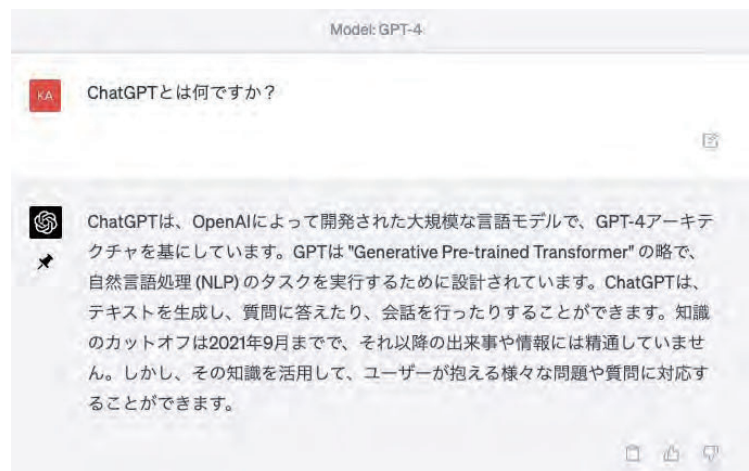


午前4:39 · 2022年9月26日 · Twitter for Android

https://twitter.com/kuron_nano/status/1574121450860007424

ChatGPT

- 対話に特化したAI (LLM)
 - 自然言語がUI
 - Webの集合知を学習
 - Human-in-the-loop (RLHF)



46
<https://atmarkit.itmedia.co.jp/ait/articles/2212/23/news019.html>

東京工業大学の場合

学修における生成系人工知能の使用に関する本学の考え方について

いいね! 230
シェアす
ツイート

授業・履修 注意喚起

RSS

公開日: 2023.04.20

学生の皆さんへ

理事・副学長(教育担当) 井村 順一

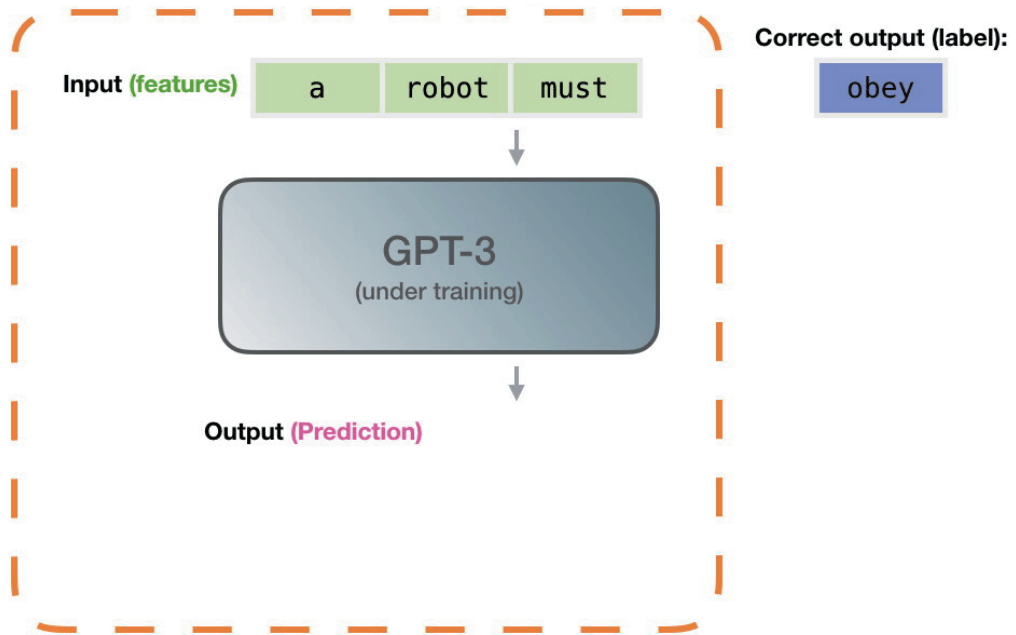
ChatGPTに代表される生成系人工知能(以下、生成系AI)が、大学におけるレポート課題などに対して使用され、世界中で大きな議論を呼んでいます。皆さんには、本学が掲げる「学生の主体的な学び=Student Centered Learning」の精神に則り、良識ある使用が求められます。本学教員とは、3月中旬に生成系AIの使用上の注意点や成績評価法について情報共有を行いました。このお知らせは、その情報並びにその後の議論を含めて、学生の皆さんに向けた形に再構成し、学修における生成系AIの使用についての本学の指針と注意点を記したものです。

本学の指針

1. 本学では、生成系AIのツールとしての危険性と有用性の両面に鑑み、現時点ではその使用を全面禁止とすることはしません。学生の皆さんの主体性を信頼し、良識と倫理観に基づいて生成系AIを道具として使いこなすことを期待します。生成系AIの出力をほぼそのまま鵜呑みにしたレポートを提出することは、皆さんが生成系AIに隷属することにも等しく、甚だ不適切です。
2. 生成系AIの使用が許される程度(禁止の有無を含む)は、授業の到達目標や内容、授業担当教員の指導方針・成績評価方針などに委ねられます。授業によっては、レポート作成やプログラミング時の利用を禁止する場合もあれば、文書校正や参考プログラムの生成などにおいて利用を推奨する場合もあるかもしれません。その詳細はシラバスに明記されていないことがありますので、授業担当教員の指示に従ってください。

<https://www.titech.ac.jp/student/students/news/2023/066590>

Unsupervised Pre-training



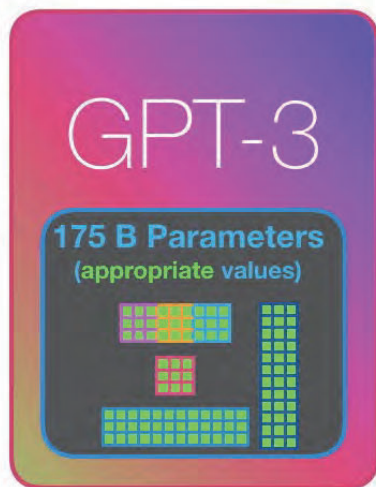
<https://jalammar.github.io/how-gpt3-works-visualizations-animations/>

Pre-training



Fine-tuning

Additional training to become better at a certain task



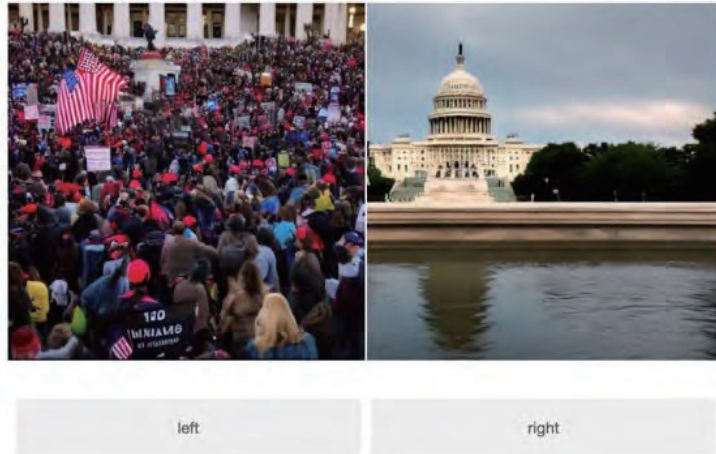
Example: English to French Translation

<https://jalammar.github.io/how-gpt3-works-visualizations-animations/>

どのようなAI生成コンテンツが拡散するか？

提示されたニュース（政治）に対して、感情価が異なる画像のうちどちらが一緒に表示された場合に、よりSNSで共有したくなるかを質問（1人12問，参加者250名）

ChatGPTで作成→ A rally against the recent election fraud is scheduled to be held again at the Capitol starting early today. (最近の選挙不正に対する抗議集会が、本日早朝から再び議事堂で開催される)



Stable Diffusionで作成→

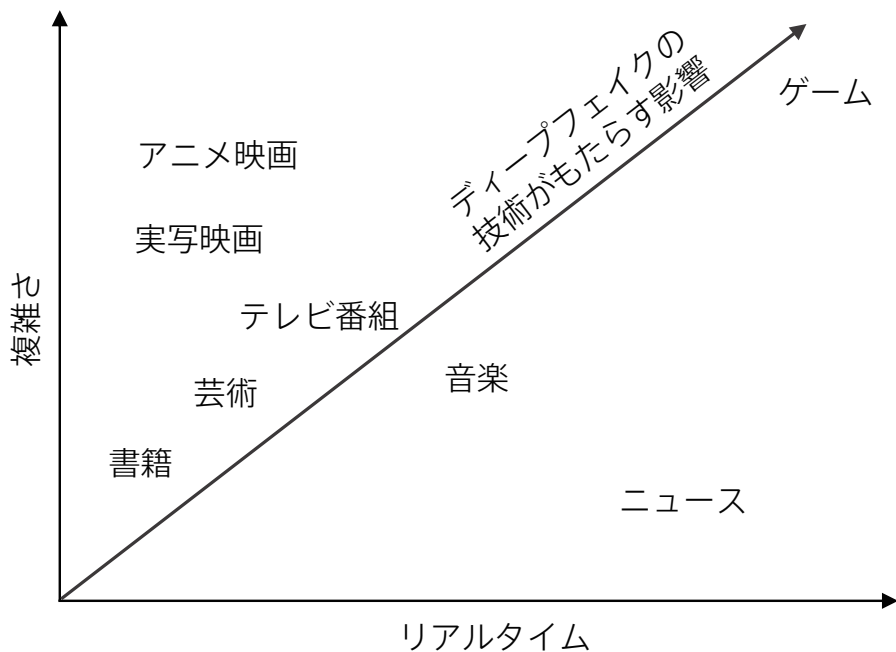
50

ネガティブな画像付きのニュースは、よりSNSで拡散される可能性がある（予備実験）

画像1と2の組み合わせ	画像1	画像2
ネガティブ, ニュートラル	2129	871
ニュートラル, ニュートラル	124	126
ネガティブ, ネガティブ	146	104

- 帰無仮説「2つの画像の選択率に差はない」でカイ二乗検定（有意水準 $\alpha=0.05$ 、ボンフェローニ補正後 $\alpha=0.017$ ）
- 「ネガティブ, ニュートラル」条件は、2つの画像の選択率に差がある→ネガティブ画像がより選択されている

ディープフェイクのもたらす影響



52

ディープフェイクの弊害

インフォカリプスへの漸近
(トラストが失われた世界)

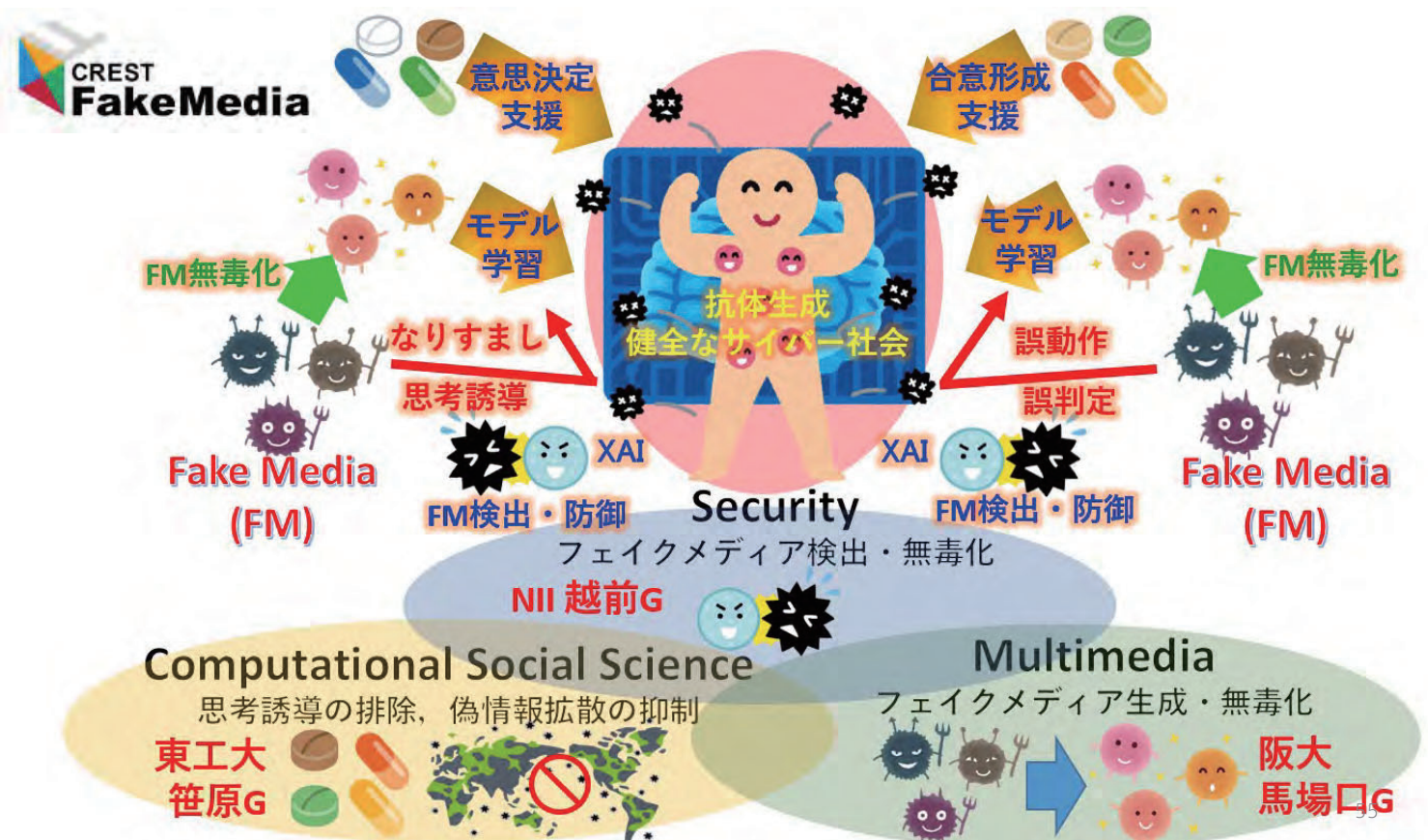
- 本物のような偽物が出回り、実害が生じる
- 本物と偽物の区別がつかない環境では、偽物を作って広めようとするものが徳をする (嘘つきの配当)
- 情報生態系の汚染

53

生成AIによる弊害＝AGCの氾濫

- 情報生態系の汚染（不確かな情報・バイアスの増幅）
- ChatGPTが生成（と共創）した情報が思考の糧に
 - “the loss of scientific diversity that may result from the removal of a barrier protecting science from homogenization” (Nakadai et al. 2023)
- ChatGPT x Stable Diffusion x SNS
 - フェイクの高度化・大量化
 - 偽のペルソナ

54



プラットフォームでの行動介入

- 情報の正確さに対する介入等 (Info Interventions)
- 多様なつながりを促進する介入 (Polyphony)

58

Info Interventions

正確性のプロンプト

虚偽の可能性のあるコンテンツに出会ったときに、ユーザに対してコンテンツの正確性に注意を向けさせると、共有する確率が減少 (次頁)

リダイレクト方式

過激派の情報を探しているユーザを、過激派の勧誘メッセージに反論する確認済みのコンテンツに自動転送

投稿者へのフィードバック

有害なコメントを検出した場合、投稿者にメッセージを表示し、公開する前に表現変更を促す

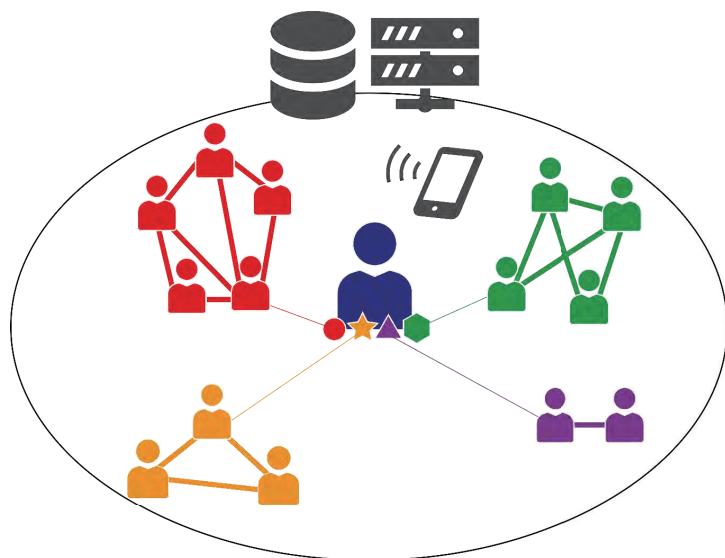
プレバンキング

デマの手口や実例を事前に知っておくことで「免疫」をつけて、実際の誤情報に出会ったときに、信じたり、拡散したりしないようにする

<https://interventions.withgoogle.com/>

59

弱い紐帯を促進するマイルドな介入



多次元の弱い紐帯の自生

- 「似てるけど少し違う」ユーザをアルゴリズムで発見
- 認知バイアスを活用してつながりを促す



社会ネットワーク全体としての
情報多様性を保全

60

Polyphony

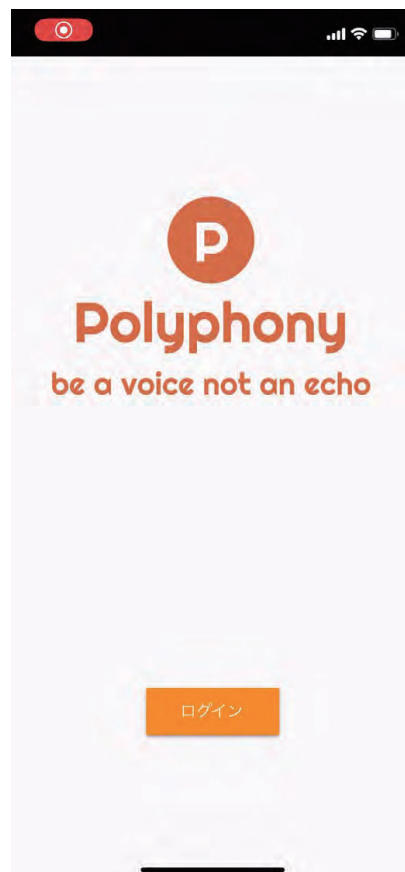
音でつながるSNS

- 投稿（ボイス）が聞ける
- 「誰かがブックマークしました」
- 「こんな人もいます」で、
相性を音で紹介（ナッジ）

ナッジ (Nudge)

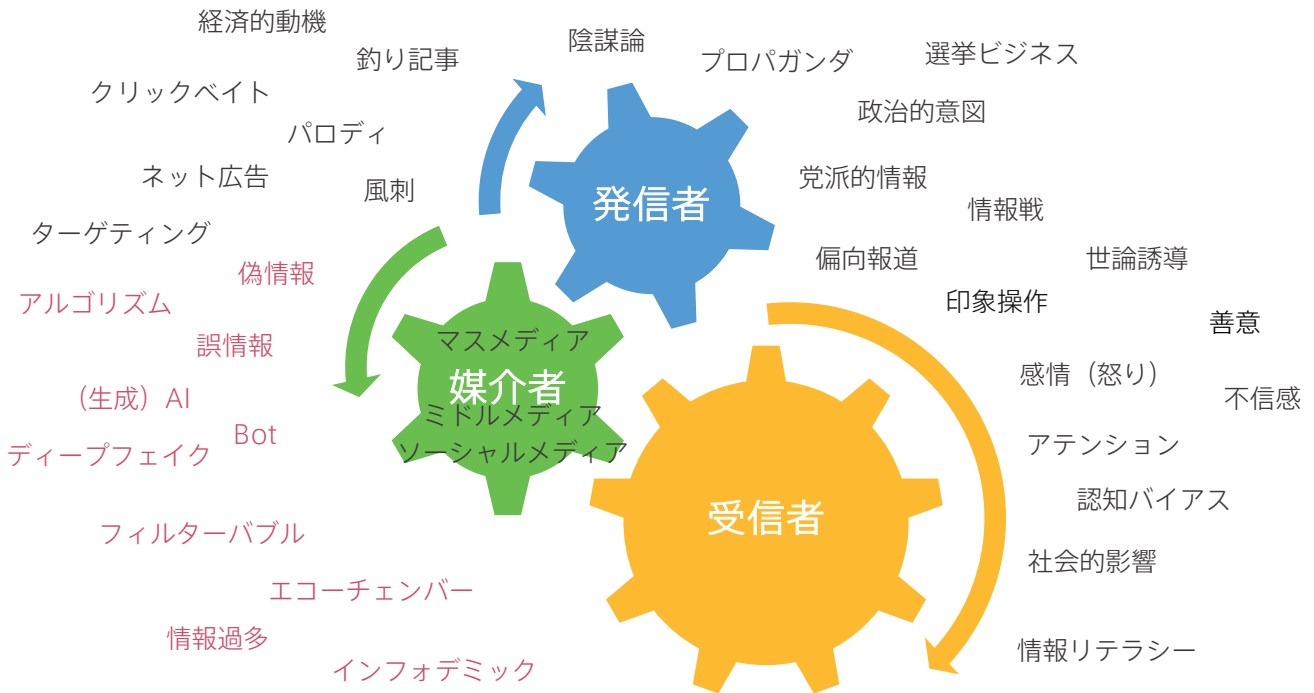
選択の余地を残しながら、自発的に
特定の行動を選択するよう促す仕掛け

笹原和俊ら, 人工知能学会全国大会論文集 (2020)

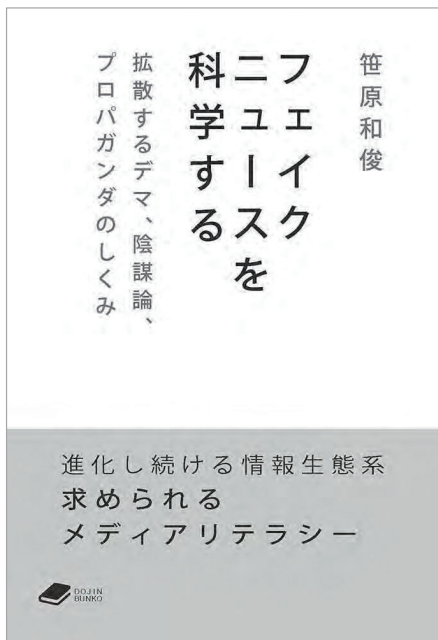


61

情報生態系の弾み車



参考文献



アドバンスソフト
からの情報提供

1. 今後のセミナー予定
2. セミナー資料のダウンロード
3. アーカイブ動画のご案内

Copyright ©2023 AdvanceSoft Corporation. All rights reserved.

アドバンスシミュレーション・セミナー2023 開催予定一覧

No.	開催日	講師の先生方	テーマ
第7回	9月14日(木) 14:00~15:30	「量子コンピュータと量子アニーリングマシン : 基礎から最先端まで」 国立研究開発法人 産業技術総合研究所 新原理コンピューティング研究センター 副研究センター長 川畑 史郎 様	量子コンピュータ
第8回	10月6日(金) 14:00~15:30	「爆轟から見える CAE の方向性」 青山学院大学 理工学部 名誉教授 林 光一 様	爆轟
第9回	11月10日(金) 14:00~15:30	「都市のデジタルツイン」 国立研究開発法人 産業技術総合研究所 人工知能研究センター 総括研究主幹 中村 良介 様	デジタルツイン
第10回	12月15日(金) 14:00~15:30	「計算物質科学による 2.5 次元物質科学研究 : 原子層物質複合構造体と外場」 筑波大学 数理物質系 教授 岡田 晋 様	ナノ

Copyright ©2023 AdvanceSoft Corporation. All rights reserved.

下記のURLから、過去のセミナー資料をダウンロード可能です。
<https://www.advancesoft.jp/download/>



シミュレーション図書館 (資料ダウンロード) について

- 初めてご利用いただくお客さまは、「新規メンバー登録 (フォーラム会員登録)」をお願いいたします。ご登録いただきますと、いただいたメールアドレスあてにダウンロード用のログインパスワードをお知らせいたします。
- ログインパスワードをお持ちのお客さまは、お客さまのご登録されたメールアドレスと、こちらからお知らせしましたパスワードで、ログインしてください。ログイン中はどの資料もダウンロードいただけます。
- 会社案内、会社概要、開発実績は、ログイン不要でダウンロードいただけます。
- ご登録いただいたメールアドレスあてに、メールマガジンやアドバンスソフトからのご案内をお送りさせていただきます場合がございますので、ご了承ください。
- 資料に含まれている内容を、その一部でも著作権者の許諾なしに、複製、改変、配布を行うことおよびインターネット上で提供する等により、一般へ送ることは法律によって固く禁止されています。

資料カテゴリー一覧

パッケージソフトウェア▷ ナノ・バイオ 半導体 流体 管路系 二相流 防災 構造 統合環境
 解析・コンサルティング▷ 防災・地震 二相流 2次電池 その他

半導体

【半導体】> セミナー資料 (アドバンスソフトの半導体デバイス、電磁波解析のご紹介 ~DX、グリーンイノベーションを加速する数値解析シミュレーション~) 20221117_all 【全体】

カテゴリー: 半導体 パッケージソフトウェア

【半導体】> セミナー資料 (半導体デバイス3次元TCADシステム Advance/TCAD 最新動向セミナー) 20181225_all 【全体】

カテゴリー: 半導体 パッケージソフトウェア

【半導体】> セミナー資料 (半導体デバイス3次元TCADシステム Advance/TCAD 製品紹介セミナー) 20190828_all 【全体】

カテゴリー: 半導体 パッケージソフトウェア

【半導体】 Advance/TCAD > セミナー資料 (オンラインセミナー「アドバンスソフトの電磁波、半導体デバイス解析のご紹介 ~IoTを支える数値解析シミュレーション~」) 20200825_all 【全体】

カテゴリー: 半導体 パッケージソフトウェア

【半導体】 Advance/TCAD > セミナー資料 (半導体デバイス3次元TCADシステム発売のご紹介セミナー)

Copyright ©2023 AdvanceSoft Corporation. All rights reserved.

YouTubeにて、セミナーのアーカイブ動画を配信中
<https://www.youtube.com/user/advancesoft>



- 【セミナー動画】第10回 アドバンス・シミュレーション・セミナー 2023年4月21日 (金) 開催 [防災・インフラ分野でのシミュレーション] advancesoft • 0 回視聴 • 35分前
- 【セミナー動画】第10回 アドバンス・シミュレーション・セミナー 2023年1月20日 (金) 開催 [環境対策を身近に感じることができる...]
- 【セミナー動画】第9回 アドバンス・シミュレーション・セミナー 2022年12月23日 (金) 開催 [「JAMSTECの最先端地球科学」回] advancesoft • 98 回視聴 • 3か月前
- 【セミナー動画】第6回 アドバンス・シミュレーション・セミナー 2022年11月11日 (金) 開催 (動的破壊進展解析と大規模粒子「...



- 【セミナー動画】 NanoLabo/NeuralMD 新機能紹介セミナー 2023年4月20日 (木) 開催 advancesoft • 2 回視聴 • 6時間前
- 【セミナー動画】 OpenModelica 活用セミナー 2022年11月30日 (水) 開催 advancesoft • 166 回視聴 • 4か月前
- 【セミナー動画】 アドバンスソフトの半導体デバイス、電磁波解析のご紹介 ~DX、グリーンイノベーションを加速する数... advancesoft • 70 回視聴 • 4か月前
- 【セミナー動画】 NanoLabo 新機能、新製品 NeuralMD Pro ご紹介セミナー 2022年11月10日 (木) 開催

Copyright ©2023 AdvanceSoft Corporation. All rights reserved.



警告

このレポートに収録されている文章および内容については、ご自身のために役立つ用途に限定して無料配布しています。このレポートを、販売、オークション、その他の目的で利用するには、著作権者の許諾が必要になります。このレポートに含まれている内容を、その一部でも著作権者の許諾なしに、複製、改変、配布を行うことおよびインターネット上で提供する等により、一般へ送ることは法律によって固く禁止されています。