

Aprendizaje profundo en sistemas de recomendación



DENIS PARRA

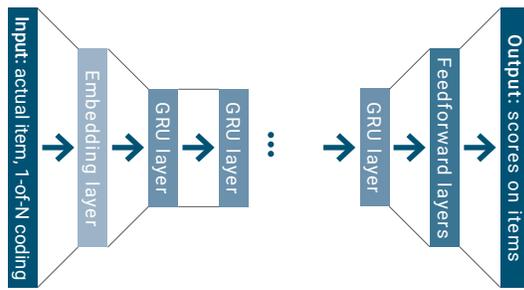
Profesor Asociado del Departamento de Ciencia de la Computación de la Pontificia Universidad Católica de Chile e Investigador Adjunto del Instituto Milenio Fundamentos de los Datos.

Corría el año 2010 y yo cursaba mi doctorado enfocado en personalización y sistemas de recomendación en la Universidad de Pittsburgh, ubicada en la ciudad homónima (Pittsburgh) al oeste del estado de Pennsylvania en Estados Unidos. Las técnicas más avanzadas de mi tema de investigación eran del área conocida como Aprendizaje Automático (en inglés, *Machine Learning*), por lo que sentía la necesidad de tomar un curso avanzado para completar mi formación. En el semestre de otoño finalmente me inscribí en el curso de Aprendizaje Automático, y gracias a un convenio académico pude cursarlo en la universidad vecina, Carnegie Mellon University. Yo estaba realmente emocionado de tomar un curso en un tema de tan creciente relevancia en unas de las mejores universidades del mundo en el área de computación.

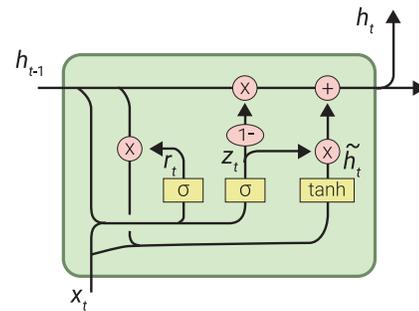
Recuerdo que vimos muchas técnicas que permitían aprender modelos a partir de datos, con especial énfasis en modelos gráficos —por ejemplo, el famoso *Latent Dirichlet Allocation* [1]— así como en métodos *kernel* como *Support Vector Machines* (SVM). Casi al final del curso, tuvimos una clase algo tímida sobre redes neuronales artificiales, un método interesante pero que poca gente usaba. Las redes neuronales artificiales datan de los años cincuenta [2], renacieron en los ochenta luego del invierno de la IA [3], para luego volver a perder tracción en los noventa. Cuál fue mi sorpresa cuando el año 2012 las redes neuronales artificiales pasaban a ser el método que todos querían usar y del cual todos hablaban. El motivo fue el sorprendente resultado del equipo SuperVision de la Universidad de Toronto¹ —Krizhevsky,

Sutskever y Hinton—, que usando una red neuronal convolucional profunda (*deep convolutional neural network*) con 60 millones de parámetros y 650 mil neuronas, entrenado con dos GPUs durante una semana, ganaba el ImageNet challenge 2012 con un error top-5 del 15,3% y más de 10 puntos de mejora en relación al segundo lugar. Las redes neuronales profundas tenían algunos antecedentes importantes de buen rendimiento [4], pero el resultado del 2012 en el ImageNet challenge catapultó su popularidad. La arquitectura de red neuronal creada empezó a ser popularmente conocida como AlexNet [5], debido al nombre del primer autor, Alex Krizhevsky. A partir de ese momento, ingenieros e investigadores de diferentes áreas de la inteligencia artificial querían escribir los términos *deep learning*

1 | <https://www.image-net.org/challenges/LSVRC/2012/results.html>.

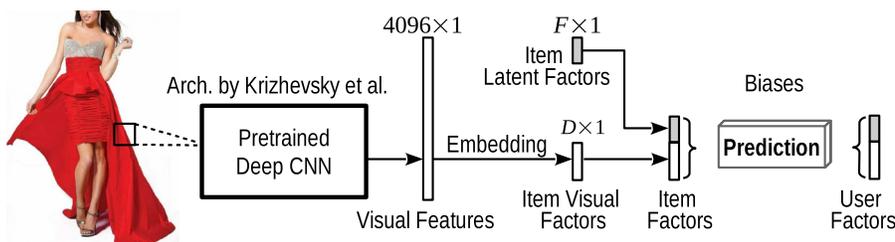


Fuente: [16].



Fuente: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>.

Figura 1. Arquitectura de GRU4Rec donde cada capa GRU tiene celdas GRU como la que se observa a la derecha, que pueden recordar y olvidar, selectivamente, permitiendo el aprendizaje de secuencias.



Fuente: [17].

Figura 2. Diagrama de VBPR que indica cómo las características visuales obtenidas con una red neuronal convolucional profunda son incorporadas en el predictor de preferencia.

o *neural network* en el título de sus artículos, y es así cómo este método empieza a permear desde el campo de visión por computador a otras áreas como recuperación de información [6], traducción automática [7], describir imágenes con texto de forma automática [8], o incluso áreas creativas como generación visual [9] y musical [10].

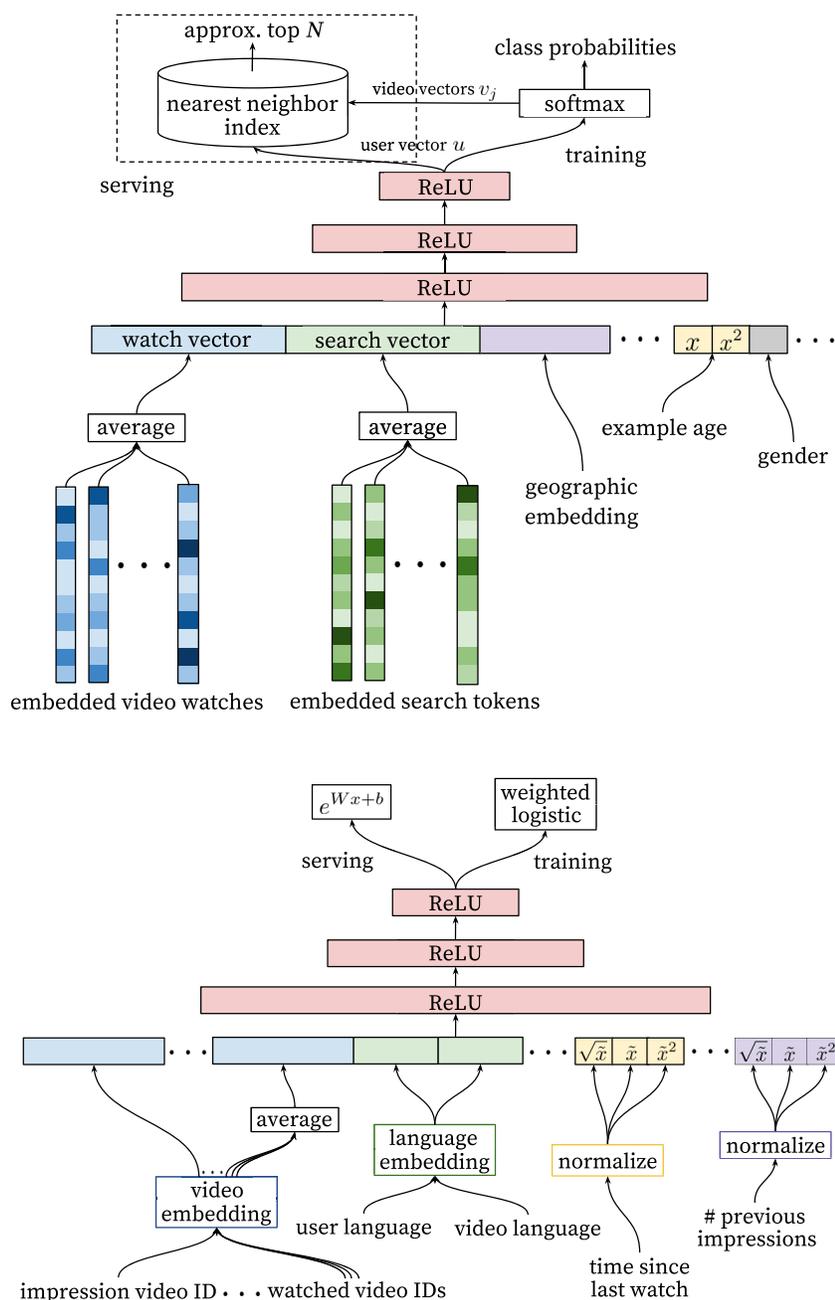
A pesar del frenesí de distintas áreas por usar aprendizaje profundo, no fue hasta el 2015 que aparecen *papers* relevantes de aprendizaje profundo aplicados a Sistemas Recomendadores (de aquí en adelante, *SisRec*). Recordemos que los *SisRec* tienen como rol principal ayudarnos a encontrar ítems relevantes dentro de una sobreabundancia de información [11] considerando nuestras

preferencias individuales. Compañías tan diversas como Amazon, Netflix, Google, Booking y Spotify basan buena parte de sus funcionalidades y modelos de negocio en sistemas recomendadores. Estos sistemas se han desarrollado por más de treinta años, pero han evolucionado especialmente rápido en la última década.

Volviendo a la aplicación de aprendizaje profundo aplicado a *SisRec*, es posible rescatar como antecedente previo a ImageNet el uso de *restricted Boltzman machines* [12], un tipo de red neuronal probabilística, entre los mejores métodos que compitieron en el Netflix prize [13]. Sin embargo, los primeros trabajos utilizando aprendizaje profundo ya sea a través de representaciones preentrenadas o para el modelo completo fueron

los trabajos de Van den Oord *et al.* [14], un recomendador de música que utilizaba representaciones de audio aprendidas con una red neuronal profunda. Luego, se presenta en 2015 “aprendizaje profundo colaborativo para *SisRec*” [15], un método que combina las técnicas de filtrado colaborativo con *denoising autoencoders*. El mismo 2015 aparece GRU4Rec [16] que modela secuencias de interacciones usando redes recurrentes con celdas GRU (ver Figura 1) para recomendar productos, y el mismo año se publica VBPR [17], método que utiliza la representación de imágenes que entrega una red convolucional preentrenada para mejorar recomendaciones visuales (ver Figura 2) realizadas por el modelo BPR [18].

Es difícil saber por qué el área de *SisRec* demoró tanto (alrededor de tres años) en ingresar a la ola de las redes neuronales profundas, pero es posible argumentar algunas razones en base a los pilares que posibilitaron el crecimiento del aprendizaje profundo: (a) gran cantidad de datos, (b) algoritmos de aprendizaje más eficientes, y (c) hardware especializado para el entrenamiento. En el área de sistemas de recomendación no era trivial encontrar *datasets* de gran tamaño, como el ImageNet, para entrenar modelos con tantos millones de parámetros como una red neuronal profunda. Esto se debe a que las grandes



Fuente: [26].

Figura 3. Las dos redes neuronales que formaban parte del sistema recomendador de videos, de aprendizaje profundo, del portal YouTube, activo hasta el 2019.

compañías han sido reticentes a compartir *datasets* que indiquen preferencias de usuarios por productos, ya sea por temas de competencia como para evitar violaciones de privacidad [19]. En los últimos años la disponibilidad de grandes *datasets* para entrenar modelos de recomendación ha mejorado mucho, con *datasets* como el de Spotify², Goodreads³ o la versión 25M del tradicional movielens dataset⁴. En cuanto a algoritmos, si bien es posible adaptar métodos existentes de clasificación de imágenes o ranking de documentos para tareas de recomendación, el hecho de tener que incorporar el modelo de usuario en el método complejiza un poco su modelamiento e implementación. No es lo mismo usar un modelo de ranking de imágenes dada una imagen de entrada, que un modelo de ranking de imágenes personalizado, que considere tanto el historial de consumo de un usuario [17, 20, 21] así como el contexto de dicho consumo —día de la semana, hora, haciendo qué actividad, etc. [22]. En relación a hardware, no es un secreto que son grandes compañías como NVidia, Google, Amazon, o Facebook quienes disponen de los mejores recursos de hardware para entrenar modelos que crecen sin cesar en cantidad de parámetros: como muestra, el reciente modelo de lenguaje GPT-3 tiene 175 mil millones de parámetros [23], comparado con los 60 millones de parámetros de la AlexNet. Esto dificulta la investigación que provenga exclusivamente desde la academia, donde los incentivos permiten investigar temas diferentes a los que empujan la investigación en la industria. A pesar de estas dificultades, una propiedad interesante de estos modelos es la posibilidad de hacer *transfer learning* [24], es decir, entrenarlos inicialmente para una tarea y luego actualizar todos o parte de sus pesos para otro *dataset* o para otras tareas. Esto permite que el costo mayor de entrenamiento lo lleven a cabo grandes compañías, fundaciones y universidades,

2 | <https://www.aicrowd.com/challenges/spotify-million-playlist-dataset-challenge>.
 3 | <https://sites.google.com/eng.ucsd.edu/ucsdbookgraph/home>.
 4 | <https://grouplens.org/datasets/movielens/25m/>.

y luego otros usuarios con menores recursos de hardware tienen sólo que adaptar (*finetuning*) los pesos para la nueva tarea o *dataset* que se aborda.

A partir del año 2016 el aprendizaje profundo aterriza con fuerza en la conferencia internacional ACM de sistemas recomendadores, donde se publica “Ask the GRU” [25], un recomendador con aprendizaje multitarea de artículos científicos que usa una red recurrente con celdas del tipo Gated Recurrent Unit. Además de este *paper*, autores de Google [26] presentan la nueva versión del sistema recomendador de videos de YouTube, basado en dos redes neuronales profundas (ver Figura 3), una red que selecciona cientos de candidatos a partir de millones de opciones, y una segunda red que ordena los videos candidatos previamente filtrados. La nueva arquitectura del portal YouTube [27] tiene algunos aspectos interesantes, por ejemplo que considera los likes de los usuarios para generar el perfil del usuario para recomendar, cosa que no hacía el recomendador anterior [26].

Luego de estas publicaciones, es común encontrar SisRec implementados con métodos de aprendizaje profundo en temas como recomendación de música, películas, libros, pareja sentimental, ropa de temporada, entre muchos otros. Los sistemas han evolucionado en los últimos años de la arquitecturas como Transformer [28], integrados con otras técnicas como aprendizaje reforzado profundo [29], así como explotando avances en áreas como NLP [30] o modelos generativos [31].

Discusión y conclusión

El aprendizaje profundo ha impactado positivamente el área de SisRec, tanto

como a otras áreas de aplicación de la inteligencia artificial. Hay, sin embargo, dos aspectos importantes a mencionar que generan inquietud en el área: cuánto es el progreso real que ha traído el aprendizaje profundo, y cómo estos modelos afectan el avance en temas de equidad, explicabilidad y transparencia.⁵

¿Cuánto se ha progresado? El artículo de [32] pone en entredicho el impacto del aprendizaje profundo en los SisRec, mostrando que cuando métodos tradicionales de factorización matricial que se conocen por más de una década son entrenados adecuadamente, tienen tanto o mejor rendimiento que métodos de aprendizaje profundo. Si bien este *paper* es relevante por mostrar una crisis de reproducibilidad en SisRec y que no siempre el aprendizaje profundo puede mejorar el rendimiento los métodos ya conocidos, hay un aspecto relevante a considerar. La investigación de Dacrema sólo considera tuplas usuario-ítem como entrada, pero no considera información adicional como imágenes, video, metadatos, contexto, etc. Justamente es con esta gran cantidad y diversidad de datos donde es esperable el rendimiento mejorado de técnicas de aprendizaje profundo, por lo cual se recomienda revisar con cautela los resultados de este análisis, y ponerlo en perspectiva sólo para el filtrado colaborativo tradicional.

FaccT. Considerar los desafíos que se plantean en la inteligencia artificial en relación a equidad (*fairness*), explicabilidad (*accountability*) y transparencia (*transparency*) es un gran desafío para los modelos de aprendizaje profundo en SisRec [33]. Considere el caso en que usa GPT-3, un modelo de 175 mil millones de parámetros, para recomendar un documento y el usuario solicita una explicación sobre dicha sugerencia ¿cómo explicaría dicha recomendación inten-

tando ser transparente? Los métodos de explicabilidad para inteligencia artificial están en activa investigación en estos días [34] y si deseamos que los sistemas de recomendación permeen áreas críticas de toma de decisiones como medicina, finanzas o seguridad, se debe avanzar en esta área. En relación a asegurar que estos sistemas no están sesgados existe una inquietud similar: cómo hacer que provean recomendaciones justas a diferentes grupo de usuarios finales, por ejemplo de un sistema de recomendación de empleo, así como a creadores de contenido: que un portal de libros recomiende con la misma probabilidad tanto a escritores hombres como mujeres o de otros grupos LGBTQ.

Conclusión. El aprendizaje profundo tomó algunos años en permear el área de sistemas de recomendación en comparación con otras áreas de inteligencia artificial, pero se instaló con fuerza a partir de 2016 gracias a su gran capacidad para encontrar representaciones de usuarios y datos para posteriormente ser usadas en tareas de filtrado de información. Con el avance de modelos de visión por computador, modelos de lenguaje, arquitecturas como atención y más recientemente modelos de redes neuronales para grafos, el impacto de las redes neuronales profundas en SisRec no ha dejado de crecer. La integración de estas técnicas con otras como aprendizaje reforzado para SisRec y el crecimiento en los últimos años de los sistemas de recomendación conversacionales [35] le siguen dando fuerza a esta área de investigación. Los desafíos en términos de mostrar los avances reales en rendimiento de estas técnicas [32] así como su adaptación para lidiar con necesidades de equidad, transparencia, explicabilidad [33], nos harán ver sin duda mucha más investigación en este tema en los años venideros. ■

5 | FaccT 2018. ACM Conference on Fairness, Accountability, and Transparency <https://faccconference.org/>.



REFERENCIAS

- [1] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research*, 3, 993-1022.
- [2] Rosenblatt, F. (1957). *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory.
- [3] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1985). *Learning internal representations by error propagation*. California Univ. San Diego La Jolla Inst. for Cognitive Science.
- [4] Ciresan, D. C., Meier, U., Masci, J., Gambardella, L. M., & Schmidhuber, J. (2011). Flexible, high performance convolutional neural networks for image classification. In *Twenty-second international joint conference on artificial intelligence*.
- [5] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 1097-1105.
- [6] Severyn, A., & Moschitti, A. (2015). Learning to rank short text pairs with convolutional deep neural networks. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval* (pp. 373-382).
- [7] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
- [8] Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2015). Show and tell: A neural image caption generator. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3156-3164).
- [9] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial networks. arXiv preprint arXiv:1406.2661.
- [10] Roberts, A., Engel, J., Raffel, C., Hawthorne, C., & Eck, D. (2018). A hierarchical latent vector model for learning long-term structure in music. In *International Conference on Machine Learning* (pp. 4364-4373). PMLR.
- [11] McNee, S. M., Kapoor, N., & Konstan, J. A. (2006). Don't look stupid: avoiding pitfalls when recommending research papers. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work* (pp. 171-180). ACM.
- [12] Salakhutdinov, R., Mnih, A., & Hinton, G. (2007). Restricted Boltzmann machines for collaborative filtering. In *Proceedings of the 24th international conference on Machine learning* (pp. 791-798).
- [13] Bennett, J., & Lanning, S. (2007, August). The Netflix Prize. In *Proceedings of KDD cup and workshop* (Vol. 2007, p. 35).
- [14] Van Den Oord, A., Dieleman, S., & Schrauwen, B. (2013). Deep content-based music recommendation. In *Neural Information Processing Systems Conference (NIPS 2013)* (Vol. 26). Neural Information Processing Systems Foundation (NIPS).
- [15] Wang, H., Wang, N., & Yeung, D. Y. (2015). Collaborative deep learning for recommender systems. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1235-1244).
- [16] Hidasi, B., Karatzoglou, A., Baltrunas, L., & Tikk, D. (2015). Session-based recommendations with recurrent neural networks. arXiv preprint arXiv:1511.06939.
- [17] He, R., & McAuley, J. (2016). VBPR: visual bayesian personalized ranking from implicit feedback. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 30, No. 1).
- [18] Rendle, S., Freudenthaler, C., Gantner, Z., & Schmidt-Thieme, L. (2012). BPR: Bayesian personalized ranking from implicit feedback. arXiv preprint arXiv:1205.2618.
- [19] Narayanan, A., & Shmatikov, V. (2006). How to break anonymity of the Netflix Prize dataset. arXiv preprint cs/0610105.
- [20] Chen, J., Zhang, H., He, X., Nie, L., Liu, W., & Chua, T. S. (2017). Attentive collaborative filtering: Multimedia recommendation with item-and component-level attention. In *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval* (pp. 335-344).
- [21] Messina, P., Domínguez, V., Parra, D., Trattner, C., & Soto, A. (2019). Content-based artwork recommendation: integrating painting metadata with neural and manually-engineered visual features. *User Modeling and User-Adapted Interaction*, 29(2), 251-290.
- [22] Adomavicius, G., & Tuzhilin, A. (2011). Context-aware recommender systems. In *Recommender systems handbook* (pp. 217-253). Springer, Boston, MA.
- [23] Brown, T. B., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165
- [24] Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10), 1345-1359.
- [25] Bansal, T., Belanger, D., & McCallum, A. (2016). Ask the gru: Multi-task learning for deep text recommendations. In *proceedings of the 10th ACM Conference on Recommender Systems* (pp. 107-114).

- [26] Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for YouTube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems* (pp. 191-198). ACM.
- [27] Zhao, Z., Hong, L., Wei, L. et al. (2019). Recommending what video to watch next: a multitask ranking system. In *Proceedings of the 13th ACM Conference on Recommender Systems* (pp. 43-51).
- [28] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. arXiv preprint arXiv:1706.03762.
- [29] Zheng, G., Zhang, F., Zheng, Z., Xiang, Y., Yuan, N. J., Xie, X., & Li, Z. (2018, April). DRN: A deep reinforcement learning framework for news recommendation. In *Proceedings of the 2018 World Wide Web Conference* (pp. 167-176).
- [30] Penha, G., & Hauff, C. (2020). What does BERT know about books, movies and music? Probing BERT for Conversational Recommendation. In *Fourteenth ACM Conference on Recommender Systems* (pp. 388-397).
- [31] Kang, W. C., Fang, C., Wang, Z., & McAuley, J. (2017). Visually-aware fashion recommendation and design with generative image models. In *2017 IEEE International Conference on Data Mining (ICDM)* (pp. 207-216). IEEE.
- [32] Dacrema, M. F., Cremonesi, P., & Jannach, D. (2019). Are we really making much progress? A worrying analysis of recent neural recommendation approaches. In *Proceedings of the 13th ACM Conference on Recommender Systems* (pp. 101-109).
- [33] Ekstrand, M. D., & Sharma, A. (2017). FATREC Workshop on Responsible Recommendation. In *Proceedings of the Eleventh ACM Conference on Recommender Systems* (pp. 382-383).
- [34] Gunning, D. (2017). Explainable artificial intelligence (xai). Defense Advanced Research Projects Agency (DARPA), nd Web, 2(2).
- [35] Christakopoulou, K., Radlinski, F., & Hofmann, K. (2016). Towards conversational recommender systems. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 815-824).