SOXTA VA HAQIQIY XABARLARNI ANIQLASHDA ISHLATILADIGAN USULLAR TAHLILI

Muhammadiyeva Dildora Kabilovna (Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti, ATDT kafedrasi professori)
Uzoqov Lochinbek Mamurjon o'g'li (Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti, doktorant)

Annotatsiya: Soxta yangiliklarni aniqlash bugungi kunda raqamli axborot oqimi ko'payishi bilan dolzarb masalaga aylanib bormoqda. Ushbu maqolada soxta va haqiqiy yangiliklarni ajratish uchun foydalaniladigan turli matn tahlil qilish usullari va zamonaviy algoritmlar tahlil qilinadi. **So'z chastotasi**, **stilometrik tahlil**, **N-gram modeli** va **sentiment tahlili** kabi an'anaviy yondashuvlar orqali matnlarning statistik va stilistik xususiyatlari o'rganiladi. Shuningdek, so'nggi texnologiyalar — **BERT**, **RoBERTa** kabi **transformer modellar** va **deep learning** algoritmlarida ishlatiladigan **attention mexanizmi**ning soxta yangiliklarni aniqlashdagi samaradorligi yoritiladi. Maqolada ushbu usullarni birgalikda qo'llash orqali soxta yangiliklarni aniqlashda yuqori samaradorlikka erishish mumkinligi muhokama qilinadi.

Kalit so'zlar: Fake yangiliklar, so'z chastotasi, sentiment tahlili, N-gram, mashinali o'qitish, transformer modellar, BERT, Deep learning.

Kirish

Soxta yangiliklar (fake news) zamonaviy raqamli dunyoning eng katta muammolaridan biri bo'lib qolmoqda. Internet, ijtimoiy tarmoqlar va bloglar orqali yangiliklar tarqalishining tezligi insonlar orasida noto'g'ri ma'lumotning keng tarqalishiga olib kelmoqda. Soxta yangiliklar nafaqat alohida shaxslarni yoki tashkilotlarni yolg'on axborot bilan adashtiradi, balki keng miqyosdagi ijtimoiy, siyosiy va iqtisodiy muammolarni keltirib chiqaradi.

Bugungi kunda, yangiliklarni aniqlash va ularning haqqoniyligini tekshirishda mashina o'rganish (Machine Learning) va tabiiy tilni qayta ishlash (Natural Language Processing, NLP) texnologiyalari katta ahamiyatga ega. Bu texnologiyalar orqali yangilik matnlarini tahlil qilish va ularning haqiqiy yoki soxta ekanligini aniqlash jarayonlari avtomatlashtirilmoqda. Matn tahlili so'nggi yillarda keng rivojlanib, turli algoritmlar va modellar yordamida soxta yangiliklarni ajratish yanada samarali amalga oshirilmoqda.

Ushbu maqolada matnlarni tahlil qilish orqali soxta va haqiqiy yangiliklarni aniqlashning samarali usullari ko'rib chiqiladi. Jumladan, so'z chastotasi, stilometrik tahlil, sentiment tahlili kabi texnikalar, shuningdek, BERT va boshqa transformer modellaridan foydalanish usullari muhokama qilinadi. Ushbu yondashuvlar soxta yangiliklarni aniqlashda qanday yordam berishi va kelajakda yanada samarali texnologiyalarni yaratish imkoniyatlari yoritiladi.

1. So'z chastotasi va stilometrik tahlil

So‘z chastotasi:

So‘z chastotasi matndagi so‘zlarning qanday va qancha marta takrorlanishini o‘rganadi. So‘zlar qanchalik tez-tez ishlatilsa, matnning mazmuni, stili yoki xabarning xususiyatlarini aniqlashga yordam beradi. Fake yangiliklarda ko‘pincha sansatsion yoki hissiy so‘zlar tez-tez takrorlanadi. So‘z chastotasi usulida ma'lum bir matndagi so‘zlarni sanash orqali tahlil qilinadi.

So‘z chastotasi formulasi:

$$TF(w) = \frac{f(w)}{N}$$

Bu yerda:

- ***TF(w)*** – matndagi so‘zning chastotasi (term frequency),
- ***f(w)*** – matnda uchragan so‘zning umumiy takrorlanish soni,
- ***N*** – matndagi jami so‘zlar soni.

Stilometrik tahlil:

Stilometrik tahlil matndagi so‘zlar, gap tuzilmalari va tilning stilistik xususiyatlarini o‘rganadi. Fake yangiliklar odatda o‘ziga xos, shov-shuvli, hissiy yoki dramatik tilda yoziladi. Stilometrik tahlil orqali quyidagi elementlarni o‘rganish mumkin:

- So‘z uzunligi va gap tuzilmasi,
- Punktuatsiyaning ishlatilishi (nuqta, undov va so’roq belgilarining soni),
- O‘zgacha so‘z birikmalarining ko‘p ishlatilishi (sensatsion yoki hissiy ifodalar).

Stilometrik tahlilning asosiy yondashuvi:

Har bir muallif yoki manbaning o‘ziga xos stilistik qo‘llanmalari bo‘ladi, bu esa fake yoki haqiqiy yangiliklarni ajratishga yordam beradi. Stilometrik tahlilni turli statistik metodlar yordamida amalga oshirish mumkin. Masalan, chi-square test orqali ikki turdagi matnlarning stilistik xususiyatlari taqqoslanadi.

2. N-gram modeli

N-gram modeli matndagi so‘zlar yoki belgilarning ketma-ketligini o‘rganish usulidir. **N** qiymati so‘zlar ketma-ketligining necha so‘zdan tashkil topishini ko‘rsatadi:

- **Unigram (N=1)**: Matndagi individual so‘zlarni o‘rganadi.
- **Bigram (N=2)**: Ikkita qo‘shni so‘zlarni o‘rganadi.
- **Trigram (N=3)**: Uchta qo‘shni so‘zlarni o‘rganadi.

Fake yangiliklar ko‘pincha hissiyotlarga boy, shov-shuvli yoki o‘ziga xos so‘z birikmalaridan foydalanadi, shuning uchun N-gram tahlili so‘zlar orasidagi bog‘liqliklarni va ushbu xabarlarning qanchalik soxta ekanligini aniqlashda yordam beradi.

N-gramning ehtimollik formulasi:

$$P(w_1, w_2, \ldots, w_n) = P(w_1) \times P(w_2 \mid w_1) \times \cdots \times P(w_n \mid w_1, \ldots, w_{n-1})$$

Bu yerda $P(w_n \mid w_1, \ldots, w_{n-1})$ — oldingi so‘zlar asosida yangi so‘zning kelish ehtimoli.

Masalan, **Bigram** uchun modelning ishlashi quyidagicha:

$$P(w_1, w_2) = P(w_1) \times P(w_2 \mid w_1)$$

Bu ko‘rsatkich orqali so‘zlar orasidagi bog‘liqlik va birikmalar qanchalik keng tarqalganligini baholash mumkin.

N-gram diagrammasi:

```
Kiruvchi matn: "Soxta yangiliklar tezda tarqaladi"

Unigrams: [Soxta], [yangiliklar], [tezda], [tarqaladi]
Bigrams: [Soxta yangiliklar], [yangiliklar tezda], [tezda tarqaladi]
Trigrams: [Soxta yangiliklar tezda], [yangiliklar tezda tarqaladi]
```

N-gram modeli ko‘p uchraydigan o’ziga xos so‘z birikmalarini va sansatsion iboralarni aniqlash uchun juda qulay.

3. Sentiment tahlili

Sentiment tahlili — bu matnning ohangi va hissiyotlarini aniqlash usuli bo‘lib, u soxta yangiliklarni aniqlashda yordam beradi. Ko‘pincha soxta yangiliklar salbiy, qo‘rquvga soluvchi yoki hissiyotlarga boy bo‘ladi, shuning uchun sentiment tahlili yordamida ularning umumiy ohangini aniqlash mumkin.

Sentiment tahlili orqali matnning **ijobiy**, **salbiy** yoki **neytral** ohangga ega ekanligini aniqlash mumkin. Ushbu tahlilda **Ohang** va **Subyektivlik** ko‘rsatkichlari muhimdir:

- **Ohang** – matnning hissiy salmog‘i (ijobiy yoki salbiy),
- **Subyektivlik** – matnning subyektiv yoki obyektiv ekanligini aniqlash.

Sentiment tahlili formulasi: Sentiment tahlilida matnning ohangi o‘zgaruvchan qiymatlar orqali baholanadi:

$$Sentiment\ score = \sum_{i=1}^{n} S(w_i)$$

Bu yerda:

- $S(w_i)$ — so‘zning ohang qiymati (ijobiy, salbiy yoki neytral),
- **n** — matndagi so‘zlar soni.

Sentiment tahlilining grafik ko‘rinishi:

Matn	Ohang	Subyektivlik
“Bu ajoyib yangilik!”	0.8 (Ijobiy)	0.6 (Subyektiv)
“Bu dahshatli!”	-0.9 (Salbiy)	0.9 (Subyektiv)

Ohang:

- Ijobiy = 0.5 dan yuqori.
- Salbiy = 0.0 dan past.
- Neytral = 0.0 dan 0.5 oralig‘ida.

Sentiment tahlili yordamida hissiyotlarga asoslangan matnlarni aniqlash va fake yangiliklarni ajratish imkoniyati oshadi. Ayniqsa, so‘z birikmalarida hissiyotlar bilan boyitilgan iboralar soxta yangiliklarning asosiy ko‘rsatkichlaridan biridir. So‘z chastotasi va stilometrik tahlil, N-gram modeli va sentiment tahlili fake yangiliklarni aniqlashda samarali usullardir. So‘z chastotasi va stilometrik tahlil matnning stilistik va statistik xususiyatlarini o‘rganishga yordam beradi. N-gram modeli esa so‘zlar orasidagi bog‘lanishlarni chuqur tahlil qiladi. Sentiment tahlili esa yangiliklardagi hissiy yondashuvlarni aniqlashda ishlatiladi. Ushbu usullarni birgalikda qo‘llash orqali fake yangiliklarni yuqori aniqlikda topish mumkin.

4. Transformer modellar (BERT, RoBERTa)

Transformer modellar

Transformer modellar, tabiiy tilni qayta ishlash (NLP) sohasida katta yutuqlarga erishdi. **Transformer modellar** ko‘pincha **recurrent neural networks (RNN)** va **convolutional neural networks (CNN)** modellari bilan bog‘liq muammolarni hal qilish uchun yaratilgan. Bu modellar matnlarni parallel qayta ishlash va uzoq muddatli bog‘liqliklarni aniqlashda yuqori natijalarga erishadi. **Transformer** arxitekturasida eng asosiy o‘ziga xoslik — bu **attention mexanizmi** bo‘lib, u matnning ma’lum qismlariga e'tibor qaratadi. Transformer modellarni quyidagi algoritmlar orasida eng samarali deb hisoblash mumkin:

- **BERT (Bidirectional Encoder Representations from Transformers)**
- **RoBERTa (Robustly Optimized BERT Pretraining Approach)**

BERT (Bidirectional Encoder Representations from Transformers) matnni ikki yo‘nalishda (bidirectional) o‘rganadigan modeldir. BERT matnning oldingi va keyingi qismlarini bir vaqtning o‘zida o‘qib, so‘zlarning aniq kontekstini tushunadi. Bu model fake yangiliklarni aniqlash uchun juda samarali, chunki u matndagi chuqur bog‘liqliklarni tushuna oladi va murakkab, kontekstual xususiyatlarni ajratib beradi. BERT modellarida oldindan o‘rgatilgan model qo‘llaniladi va u ma'lum bir soha uchun qayta o‘qitilishi mumkin, bu esa fake yangiliklarni aniqlashda maxsus tayyorlangan model yaratish imkoniyatini beradi.

RoBERTa (Robustly Optimized BERT Pretraining Approach) — BERT modelining optimallashtirilgan versiyasidir. RoBERTa BERT bilan solishtirganda quyidagi takomillashtirishlarga ega:

- Matnni qayta o‘rganish uchun ko‘proq ma’lumotdan foydalanadi.
- Dynamic masking texnikasi bilan xususiyatlarni yaxshiroq o‘rganadi.
- O‘quv davrida ko‘proq iteratsiyalarni amalga oshiradi.

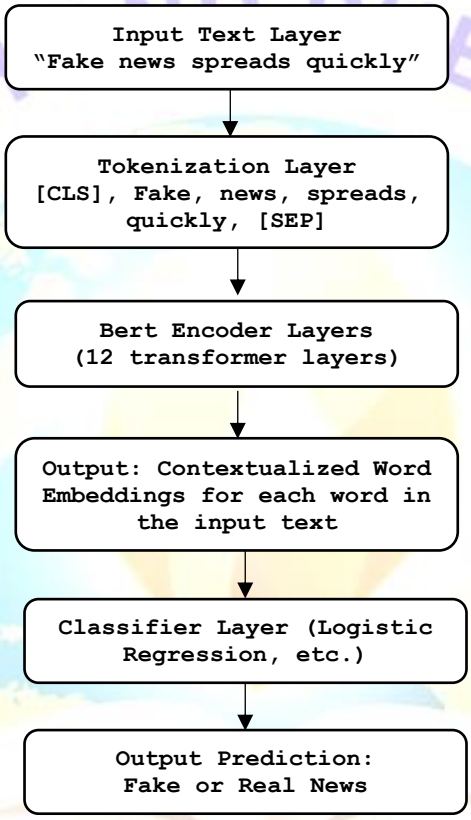
Bu modellar fake yangiliklarni aniqlashda chuqur kontekstual bog‘liqliklarni tahlil qiladi va matnning chuqur tahlilini ta’minlaydi.

BERT modeli ishlash jarayoni (bloki):

BERT modelining asosiy vazifasi matnning ikki tomonlama (bidirectional) kontekstini o‘rganishdir. Quyidagi diagramma BERT modelining ishlash jarayonini ko‘rsatadi:

BERT model diagrammasi:

- **Input Text Layer**: Bu yerda matn kiritiladi. BERT modelida har bir so‘z kontekstini tushunish uchun “[CLS]” boshlang‘ich belgisi va “[SEP]” ajratuvchi belgilari ishlatiladi.



- **BERT Encoder Layers**: Matn transformer qatlamlari orqali o‘tkaziladi va har bir so‘z uchun kontekstual tushuncha hosil qilinadi. Bu qatlamlar matnni ikki yo‘nalishda (bidirectional) o‘rganadi.
- **Classifier Layer**: BERT tomonidan chiqarilgan xususiyatlar tasniflash qatlami (masalan, Logistic Regression) orqali soxta yoki haqiqiy deb belgilangan natijaga aylanadi.

5. Attention mexanizmi

Attention mexanizmi Deep Learning modellariga qaysi so‘zlar yoki gaplar muhimroq ekanligini belgilash imkonini beradi. Matnning muhim qismlariga e'tibor qaratib, model fake va haqiqiy yangiliklarni aniqroq ajrata oladi. Attention mexanizmi BERT va transformer modellarida keng qo‘llaniladi.

Attention mexanizmining asosiy yondashuvi:

- Matndagi har bir so‘z uchun uning boshqa so‘zlar bilan bog‘liqligi va ahamiyatini o‘lchaydi.

- Model har bir so‘z uchun e'tiborni qanday taqsimlash kerakligini o‘rganadi.

Self-Attention mexanizmi

Self-attention — bu transformer arxitekturasining markaziy elementi bo‘lib, u matndagi har bir so‘zning boshqa so‘zlarga bo‘lgan ta'sirini o‘lchaydi. Self-attention mexanizmi yordamida model uzoq masofadagi bog‘lanishlarni hisobga oladi.

Self-attention ishlash jarayoni:

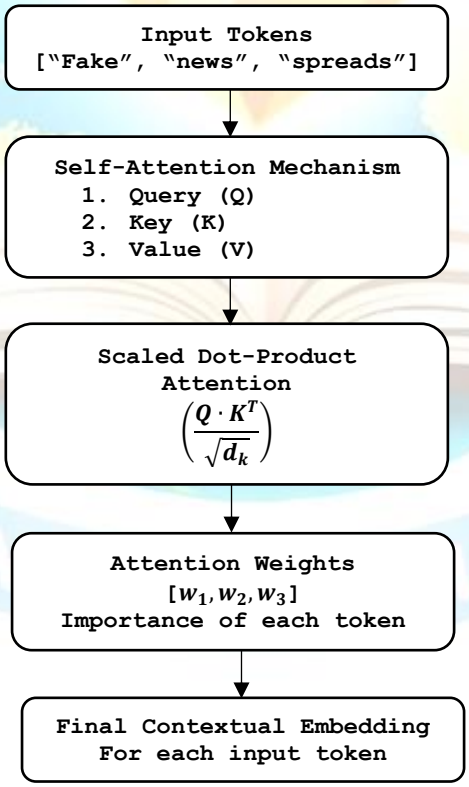
$$Attention(Q,K,V) = softmax\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right) \cdot V$$

Bu yerda:

- **Q (Queries)** — hozirgi so‘zning bog‘liqligini so‘raydi.
- **K (Keys)** — boshqa so‘zlarning bog‘liqlik koeffitsiyentini ta'minlaydi.
- **V (Values)** — so‘zning bog‘liqlik darajasini belgilaydi.
- d_k — normalizatsiya ko‘rsatkichidir.

Attention mexanizmi bilan ishlaydigan Deep Learning modellar fake yangiliklarni aniqlashda sezilarli natijalar beradi, chunki ular nafaqat har bir so‘zning ma'nosini, balki so‘zlar orasidagi kontekstual bog‘liqlikni ham tushunadi.

Self-Attention mexanizmi diagrammasi:



- **Query (Q)**, **Key (K)**, va **Value (V)** matnning har bir so‘zi uchun o‘rnatiladi.
- So‘zlar orasidagi bog‘liqlik **Scaled Dot-Product Attention** orqali o‘lchanadi va **softmax** yordamida e'tibor darajasi taqsimlanadi.

- Har bir so‘zning matndagi boshqa so‘zlarga bo‘lgan ta'siri hisoblanadi va natijada har bir so‘z uchun kontekstual embeddinglar hosil qilinadi.

Attention mexanizmi va Transformer modelining ishlashi

Transformer modelida attention mexanizmi har bir qatlamda ishlatiladi, va har bir so‘zning boshqa so‘zlar bilan bog‘liqligini aniqlashda yordam beradi. Quyida transformer modelidagi attention jarayonini ko‘rsatamiz:

Transformer model diagrammasi:

1. **Self-Attention Layer**: Matndagi har bir so‘z uchun boshqa so‘zlar bilan bog‘liqlik o‘rganiladi.
2. **Feed-Forward Layer**: Attention mexanizmi orqali hisoblangan qiymatlar transformatsiya qilinadi va keyingi qatlamga uzatiladi.

Transformer modellar (BERT, RoBERTa) va **Deep Learning** modellaridagi **attention mexanizmi** fake yangiliklarni aniqlashda juda samarali texnologiyalar hisoblanadi. BERT va RoBERTa matnning kontekstual tushunchalarini aniqlashda va ular orasidagi bog‘liqliklarni o‘rganishda chuqur natijalarga erishadi. Attention mexanizmi esa modelning muhim qismlarga e'tibor qaratishini ta'minlaydi, bu esa fake yangiliklarni aniqroq tasniflashga imkon beradi. Ushbu yondashuvlar fake yangiliklarga qarshi kurashda katta ahamiyatga ega va kelajakda yanada ko‘proq takomillashtirilishi mumkin.

Xulosa

Soxta yangiliklarni aniqlash bugungi raqamli davrning muhim va dolzarb muammolaridan biri bo‘lib, bu sohada samarali va ishonchli texnologiyalarni qo‘llash zarurati ortib bormoqda. Ushbu maqolada ko‘rib chiqilgan matn tahlil qilish usullari — **so‘z chastotasi va stilometrik tahlil**, **N-gram modeli**, **sentiment tahlili**, hamda zamonaviy **transformer modellar (BERT, RoBERTa)** va **attention mexanizmlari** soxta yangiliklarni aniqlash uchun kuchli vositalar hisoblanadi.

So‘z chastotasi va **stilometrik tahlil** matnning stilistik va statistik jihatlarini tahlil qilish orqali soxta yangiliklarni ajratishga yordam beradi. **N-gram modeli** matndagi so‘zlar orasidagi bog‘liqlik va ketma-ketlikni o‘rganib, so‘zlarning qanchalik sansatsion yoki hissiy tilda qo‘llanilganini aniqlashga imkon beradi. **Sentiment tahlili** esa yangiliklarning hissiyotlar va ohangini aniqlab, ularning salbiy yoki ijobiy ekanligini tahlil qilish orqali soxta yangiliklarni aniqlashni osonlashtiradi.

So'nggi yillarda rivojlangan **transformer modellar** (masalan, BERT va RoBERTa) matnni ikki yo‘nalishda (bidirectional) o‘rganib, chuqur kontekstual tahlil qiladi. Bu modellar fake yangiliklarni aniqlashda sezilarli natijalarga erishdi, chunki ular matndagi uzoq masofadagi bog‘liqliklarni ham o‘rganadi. **Attention mexanizmi** esa matndagi

muhim qismlarga e'tibor qaratib, fake yangiliklarni aniqlash samaradorligini yanada oshiradi.

Attention mexanizmi bilan qo‘llanilgan yondashuvlar fake yangiliklarni yuqori aniqlikda ajratishga yordam beradi, chunki ular so‘zlar va jumlalar orasidagi chuqur bog‘lanishni hisobga oladi. BERT va RoBERTa modellarining yordami bilan fake yangiliklarni aniqlash texnologiyasi aniqroq va ishonchli bo‘lib bormoqda. Bu texnologiyalar ijtimoiy tarmoqlarda va internetda tarqaladigan soxta xabarlarni tezkorlik bilan aniqlashda yordam berib, global axborot xavfsizligini ta’minlashda muhim ahamiyat kasb etadi.

Kelajakda ushbu algoritmlar yanada takomillashtirilishi va real vaqt rejimida ishlaydigan tizimlarga joriy qilinishi kutilmoqda. Shuningdek, ko‘p tilli datasetlarni yaratish, multimodal tahlillar (matn, rasm, video) orqali soxta yangiliklarni aniqlash yondashuvlari yanada keng qo‘llanilishi mumkin. Shu bilan birga, modellarni izohlaydigan (Explainable AI) yondashuvlar orqali natijalarni tushunish va ishonchni oshirish mumkin bo‘ladi. Shu tarzda, fake yangiliklar muammosini yechishda mashina o‘rganish va tabiiy tilni qayta ishlash texnologiyalari katta yordam beradi.

Adabiyotlar ro‘yxati

1. **Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K.** (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT 2019*, 4171-4186. doi:10.18653/v1/N19-1423.

2. **Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., & Stoyanov, V.** (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv preprint arXiv:1907.11692*.

3. **Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I.** (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems (NeurIPS)*, 5998-6008.

4. **Pennington, J., Socher, R., & Manning, C. D.** (2014). GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532-1543. doi:10.3115/v1/D14-1162.

5. **Zhou, X., & Zafarani, R.** (2018). Fake News: A Survey of Research, Detection Methods, and Opportunities. *arXiv preprint arXiv:1812.00315*.

6. **Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H.** (2017). Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter*, 19(1), 22–36. doi:10.1145/3137597.3137600.

7. **Rubin, V. L., Conroy, N. J., Chen, Y., & Cornwell, S.** (2016). Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News. *Proceedings of the Second Workshop on Computational Approaches to Deception Detection*, 7–17. doi:10.18653/v1/W16-0802.

8. **Hochreiter, S., & Schmidhuber, J.** (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780. doi:10.1162/neco.1997.9.8.1735.

9. **Goodfellow, I., Bengio, Y., & Courville, A.** (2016). *Deep Learning*. MIT Press.

10. **Goldstein, J., & Gigerenzer, G.** (2002). Models of Ecological Rationality: The Recognition Heuristic. *Psychological Review*, 109(1), 75-90. doi:10.1037/0033-295X.109.1.75.