

Время выполнения задания – 180 мин., язык - русский.

Максимальное количество баллов – 100.

I. ОБЩАЯ ЧАСТЬ

Прочтите отрывок статьи «Five Things That Scare Me About AI» (Rachel Thomas)

(<https://www.fast.ai/2019/01/29/five-scary-things>).

Составьте краткое резюме этого отрывка на русском языке (объемом в один абзац, но не более 1000 знаков).

- а) Каков основной тезис?
- б) Какие аргументы за или против выдвигаемого тезиса приводит автор?
- в) Выскажите ваши собственные соображения по проблемам, затронутым в отрывке, и их решениям.
- г) Какие способы применения AI для автоматической обработки текста вы можете назвать?

«AI is being increasingly used to make important decisions. Many AI experts (including Jeff Dean, head of AI at Google, and Andrew Ng, founder of Coursera and [deeplearning.ai](https://www.fast.ai)) say that warnings about sentient robots are overblown, but other harms are not getting enough attention. I agree. I am an AI researcher, and I'm worried about some of the societal impacts that we're already seeing. In particular, these things scare me about AI:

1. Algorithms are often implemented without ways to address mistakes.
2. AI makes it easier to not feel responsible.
3. AI encodes & magnifies bias.

1. Algorithms are often implemented without ways to address mistakes.

After the state of Arkansas implemented software to determine people's healthcare benefits, many people saw a drastic reduction in the amount of care they received, but were given no explanation and no way to appeal. Tammy Dobbs, a woman with cerebral palsy who needs an aid to help her to get out of bed, to go to the bathroom, to get food, and more, had her hours of help suddenly reduced by 20 hours a week, transforming her life for the worse. Eventually, a lengthy court case uncovered errors in the software implementation, and Tammy's hours were restored (along with those of many others who were impacted by the errors).

Observations of 5th grade teacher Sarah Wysocki's classroom yielded positive reviews. Her assistant principal wrote, "It is a pleasure to visit a classroom in which the elements of sound teaching, motivated students and a positive learning environment are so effectively combined." Two months later, she was fired by an opaque algorithm, along with over 200 other teachers. The head of the PTA and a parent of one of Wyscoki's students described her as "One of the best teachers I've ever come in contact with. Every time I saw her, she was attentive to the children, went over their schoolwork, she took time with them and made sure." That people are losing needed healthcare without an explanation or being fired without explanation is truly dystopian!

2. AI makes it easier to not feel responsible.

Unfortunately, complex systems lend themselves to a dynamic in which nobody feels responsible for the outcome.

An example of nobody feeling responsible comes from the case of research to classify gang crime. A database of gang members assembled by the Los Angeles Police Department (and 3 other California law enforcement agencies) was found to have 42 babies who were under the age of 1 when added to the gang database (28 were said to have admitted to being gang members). Keep in mind these are just some of the most obvious errors- we don't know how many other people were falsely included. When researchers presented work on using machine learning on this data to classify gang crimes, an audience member asked about ethical concerns. "I'm just an engineer," responded one of the authors.

I don't bring this up for the primary purpose of pointing fingers or casting blame. However, a world of complex systems in which nobody feels responsible for the outcomes (which can include severely disabled people losing access to the healthcare they need, or innocent people being labeled as gang members) is not a pleasant place. Our work is almost always a small piece of a larger whole, yet a sense of responsibility is necessary to try to address and prevent negative outcomes.

3. AI encodes & magnifies bias.

Bias is being encoded and even magnified in a variety of applications:

** software used to decide prison sentences that has twice as high a false positive rate for Black defendants as for white defendants*

** computer vision software from Amazon, Microsoft, and IBM performs significantly worse on people of color*

** Word embeddings, which are a building block for language tools like Gmail's SmartReply and Google Translate, generate useful analogies such as Rome:Italy :: Madrid:Spain, as well as biased analogies such as man:computer programmer :: woman: homemaker.*

** Machine learning used in recruiting software developed at Amazon penalized applicants who attended all-women's colleges, as well as any resumes that contained the word "women's."*

** Over 2/3 of the images in ImageNet, the most studied image data set in the world, are from the Western world (USA, England, Spain, Italy, Australia).*

Since a Cambrian explosion of machine learning products is occurring, the biases that are calcified now and in the next few years may have a disproportionately huge impact for ages to come (and will be much harder to undo decades from now).

II. СПЕЦИАЛЬНАЯ ЧАСТЬ

Задание 2. Решите задачу.

В книге 112 страниц. Что написано на первых 100 из них — неизвестно. На 101-й написано: «На всех предыдущих страницах написано одно ложное утверждение» (и больше ничего). На 102-й написано: «На всех предыдущих страницах написаны два ложных утверждения» (и больше ничего) и так далее, на 112-й написано: «На всех предыдущих страницах написано 12 ложных утверждений». Среди утверждений, написанных на последних

Олимпиады для студентов и выпускников «Высшая лига» – 2020 г.

12 страницах, есть хотя бы одно истинное. Сколько всего ложных утверждений в книге (включая первые сто страниц)? Докажите.

Задание 3

Предложите обоснованное решение.

Вы хотите построить чат-бот, который мог бы поддерживать простой бытовой диалог (например, отвечать на вопросы “какие книги тебе нравятся?”, “как настроение?”, “что сегодня делал?”). Вне зависимости от того, будет чат-бот основан на правилах или на машинном обучении, вам понадобится база вопросов и ответов. Подумайте над разными сценариями реализации программы и ответьте на следующие вопросы:

- A) В какой форме можно представить вопросы и ответы для реализации, основанной на правилах, что нужно будет включить в словарь? Опишите и/или представьте схематично. Приведите пример единицы словаря. Опишите метод, которым можно получить такие единицы.
- B) Какого рода правила могут понадобиться для построения диалога?
- C) Как будет устроена база вопросов и ответов для реализации машинного обучения? Опишите и/или представьте схематично. Опишите какой-нибудь способ сбора/генерации/расширения обучающего корпуса (откуда берется, объем, другие параметры).
- D) На основе чего будет подбираться подходящий ответ на вопрос пользователя, при реализации на машинном обучении.
- E) Какие преимущества и недостатки с точки зрения результата и особенностей реализации есть у каждого из подходов?
- G) Как бы вы стали оценивать качество работы получившегося чат-бота? Какой могла бы быть система рейтингов (штрафов, баллов и т.п.)? Предложите систему (схему) оценивания.