

**Критерии выполнения заданий.**

**Задание 1 — 40 баллов**

Основной тезис: системы на базе AI становятся все более актуальны и применяются в практических задачах во многих областях человеческой жизни, а не только IT-специалистами и учеными.

Проблемы, оговариваемые в тексте: применение AI в медицине, животноводстве и использовании солнечной энергии; доступность технологий машинного обучения людям не из сферы IT.

**а) Тезис (макс. 5 баллов)**

- Если указано только применение в разных областях без сопоставления с профессионалами — 2 балла
- Если тезис выделен полностью — **5 баллов**

**б) Плюсы и минусы (макс. 5 баллов)**

- Описание аргументации автора статьи; краткое описание примеров трех случаев применения AI на практике, упомянутых в статье — **5 баллов**
- Общие фразы вроде “автор приводит в пример истории трех людей” без подробностей (хотя бы кто какую задачу решал) — **2 балла**

**в) Собственное мнение (макс. 10 баллов)**

- Собственное мнение выражено; ясность формулировок, логичность излагаемой позиции, аргументация, примеры (свои примеры или собственный анализ приведенных в статье) — **10 баллов**
- Общие фразы без конкретизации с помощью примеров или пример без мнения — **5 баллов**
- Только фраза “я (не) согласен с автором” без аргументации — **1 балл**

**г) Применение AI в обработке естественного языка (макс. 15 баллов)**

- Перечислено только упомянутое в статье — **1 балл**
- Перечислил технологии АОТ, не связанные с AI (Томита-парсер, онтологии и т.п.) — **3 балла**
- Перечислены задачи, которые могут быть решены с использованием AI (автоматический перевод, построение диалоговых систем), но без способов решения, или технологии (морфологический анализ и т.п.) без примеров использования — **5 баллов**
- Перечислены технологии АОТ, не связанные с AI (или как связанные, так и не связанные) и приведены примеры использования — **10 баллов**

- Перечислены технологии АОТ с использованием AI и приведены примеры их применения — **15 баллов**

д) Культура академического письма и стилистика работы — 5 баллов

- Текст не разбит на абзацы (резюме статьи, собственное мнение, описание технологий АОТ), но в целом написан грамотно — **2 балла**
- Ответ оформлен в виде списка пунктов, а не связного текста — **3 балла**
- Работа четко структурирована, написана грамотным академическим языком — **5 баллов**

**Задание 2 (компьютерная лингвистика) — 10 баллов**

- Неверный ответ — **0 баллов**
- Верный ответ без обоснования — **5 баллов**
- Верный ответ с обоснованием — **10 баллов** (есть описание хода решения, формулы)

**Задание 3 (компьютерная лингвистика) — 50 баллов**

1. Приведено и объяснено формальное описание метра (метрические шаблоны) — **5 баллов**

2.

а. Правильный подход (**макс. 10 баллов**)

- Акцентуация — **2 балла**
- Слоговое деление и подсчет слогов — **2 балла**
- Определение рифмы — **2 балла**
- Семантическое согласование / сочетаемость по корпусу — **2 балла**
- Примеры правил — **2 балла**

б. Морфологический и синтаксический анализ (**макс. 10 баллов**)

- Описано, зачем нужен морфологический анализатор, что он из себя представляет и как осуществляется (для разбора и генерации словоформ; примеры анализаторов — `rumorphy`, `mystem`) — **5 баллов**
- Общие фразы про необходимость морфологического анализа без описания решения этой задачи — **2 балла**
- Описано, зачем нужен синтаксический анализ, как он осуществляется, какой формализм можно использовать — **3 балла**
- Описание общей архитектуры системы (из каких компонентов состоит, в каком порядке проводится обработка текста и т.п.) — **2 балла**

### с. Машинное обучение (макс. 15 баллов)

- Назван объект — **3 балла**
- Перечислены признаки (“фичи”) — **5 баллов**
- Объяснена необходимость или необязательность синтаксических признаков, приведены примеры — **2 балла**
- Названы подходящие методы машинного обучения — **5 баллов**
- Названы методы машинного обучения одним словом, без конкретизации (например, “нейросеть”) — **2 балла**

### 3. Тестирование системы (макс. 10 баллов)

- Названы, а также подробно описаны и обоснованы критерии, по которым можно тестировать систему; представлена система рейтингов (штрафов, баллов) для разных лингвистических функций — **4 балла**
- Перечислена пара критериев без обоснования, описания — **2 балла**
- Приведены примеры оценки различных ответов системы, формулы расчета результирующей оценки, метрики — **3 балла**
- Описаны итоги тестирования (что получится на выходе? как интерпретировать результирующую оценку? как это поможет улучшить работу системы?) — **3 балла**

### Задание 2 (ДН) — 20 баллов

- Правильный ответ (рамал-и мусаддас-и махзуф) — **10 баллов**. Попадание  $\frac{2}{3}$  (например хазадж-и мусаддас-и махзуф, рамал-и мусамман-и махзуф) при правильной аргументации верной части ответа — **8 баллов**. Попадание  $\frac{1}{3}$  при правильной аргументации верной части ответа — **3 балла**.
- Правильная аргументация — **10 баллов**. Частично правильная аргументация (2 верных) — **8 баллов**. Частично правильная аргументация (1 верный) — **5 баллов**.

Переносимые из русского языка лингвистические представления» — это понятие слогоделения, которое производится так же, как и в современном русском языке. Такое вводное соображение может служить подсказкой, потому что не так много лингвистических представлений можно перенести с одного языка на другой неродственный язык.

Аруз как система стихосложения основывается на регулярном распределении долгих и кратких слогов. Но в старотатарском языке фонологическое противопоставление долготы и краткости отсутствует, поэтому адаптация аруза строится на «условно долгих» и «условно кратких» слогах. «Условно краткий» слог всегда открытый. «Условно долгий» может быть и открытым, и закрытым.

*Мусаддас-и махзуф, мусамман-и махзуф, мусамман-и салим* отличаются только длиной.

*Рамал-и мусаддас-и махзуф и хазадж-и мусаддас-и махзуф* каждый по 11 слогов. Строки, размер которых нужно определить, тоже имеют длину в 11 слогов. Поэтому нужно выбрать только между рамалом и хазаджем.

Хазадж строится по схеме:

21112111211

Здесь символ 2 означает условно краткий слог, а символ 1 условно долгий.

Таким образом, все слоги под номерами 1, 5, 9 обязательно должны быть открытыми.

Остальные слоги могут быть как открытыми, так и закрытыми. Действительно, в приведенных примерах хазаджа (любой длины) слоги 2, 3, 4, 6, 7, 8, 10, 11 имеют коллизии с точки зрения открытости/закрытости, слоги 1, 5, 9 всегда открытые:

<sup>1</sup>A<sup>2</sup>nyŋ <sup>3</sup>kem <sup>4</sup>al <sup>5</sup>a<sup>6</sup>ŋin<sup>7</sup>dä <sup>8</sup>mäŋ <sup>9</sup>ja<sup>10</sup>rat<sup>11</sup>ty

Рамал строится по схеме:

12111211121

Все слоги под номерами 2, 6, 10 обязательно открытые, остальные могут быть и открытыми, и закрытыми.

Разобьем строки из задания на слоги:

<sup>1</sup>Äs<sup>2</sup>sä<sup>3</sup>lam <sup>4</sup>wä <sup>5</sup>äs<sup>6</sup>sä<sup>7</sup>lam <sup>8</sup>wä <sup>9</sup>äs<sup>10</sup>sä<sup>11</sup>lam;  
<sup>1</sup>I <sup>2</sup>ža<sup>3</sup>nyj, <sup>4</sup>bäy<sup>5</sup>dä <sup>6</sup>sä<sup>7</sup>lam <sup>8</sup>bu<sup>9</sup>lyr <sup>10</sup>qä<sup>11</sup>lam

В первой строке открытые слоги имеют номера 2, 4, 6, 8, 10, во второй — 1, 2, 5, 6, 8, 10. Слоги, которые являются открытыми и в первой строке, и во второй, имеют номера 2, 6, 8,

10. Ни один из этих слогов не совпадает с «условно кратким» хазаджа, но три из четырех совпадают с «условно краткими» рамала.

Искомый размер — рамал-и мусаддас-и махзуф.

Решающий задачу может понять задание так, что нужно отдельно определить размер первой строки и размер второй строки, метрическая характеристика может не совпадать. Во-первых, такой трактовке условия противоречит явная рифма между этими строками ässälam — qälam, то есть что эти строки не связаны и, следовательно, у них может быть разный размер, становится маловероятным. Во-вторых, это не важно, потому что даже по отдельности номера открытых слогов первой и второй строки не совпадают с номерами слогов, которые обязательно должны быть открытыми в хазадже.

### Задание 3 (ДН) — 40 баллов

Сбор материала.

1. **10 баллов** за указание конкретных источников (stihi.ru/ proza.ru/ samlib / паблики типа «все поэты вконтакте» ) / **8 баллов** за описание источников в общем виде (“сайты любительской литературы”, “vk-сообщества любительской поэзии”)
2. **10 баллов** — упоминание конкретных технологий веб-скрейпинга в любом виде (утилита wget, прогр.библиотеки) / **5 баллов** — общее описание логики массового обкачивания источников / **1 балл** — описание ручного сбора материала из веб-источников или общие фразы (взять из онлайн-источников..)
3. Разметка. Морфологическая разметка + **2 балла**. Семантическая разметка любого типа + **2 балла**. Синтаксическая разметка + **2 балла**. Разметка

сюжетных функций/мотивов/любой другой сложной филологической категории, которую можно выделить в виде отрезка текста, + **4 балла**. Метаразметку в этом пункте не засчитываем, засчитываем в следующем.

4. Для каждого сценария: наличие сформулированной исследовательской цели + **4 балла**. Наличие выполнимых последовательных шагов + **6 баллов**. Далее усреднение оценки по всем сценариям (итого **10 баллов**)