



### 3. [RADAR] Выявление и оценка горячих новостей в финансах

**Коротко:** Система поиска и оценки новостей в финансовой сфере с генерацией черновика для поста/статьи

**Данные:** Датасет не предоставляем; внешние источники необходимы

**Оценка:** Питч + проверка работоспособности

#### Контекст

Финансовые рынки живут в режиме **информационного переизбытка**: десятки тысяч публикаций в сутки (релизы компаний, регуляторы, отраслевые СМИ, соцсети, судебные документы, ленты агентств). Ценность — в **скорости** и **материальности**: кто первым выделил значимое среди шума и связал это с конкретными активами и рисками.

**Что считаем «горячей» новостью:** неожиданность относительно консенсуса, **материальность** для цены/волатильности/ликвидности, скорость распространения (репосты, зеркала, апдейты), широта затрагиваемых активов (прямые и косвенные эффекты), достоверность источника и подтверждений.

Как «горячее» проявляется — **два базовых сценария**:

- **Мгновенный импульс:** единичное сообщение (форс-мажор, регуляторное решение, guidance компании) мгновенно меняет оценку рисков.
- **Нарастающий нарратив:** цепочка публикаций («утечка → инсайт отраслевого издания → подтверждение в релизе → уточнение

параметров») приводит к переоценке спустя несколько итераций.

#### ▼ Эффект гетерогенен

- **По активам и секторам:** чувствительность различается (капитализация, долговая нагрузка, факторная экспозиция, индексное включение).
- **По каналам и языкам:** регистратуры/регуляторы, локальные СМИ и X/Telegram дают разные сигналы; важны мульти-язык и часовые пояса.
- **По режиму рынка:** при высокой неопределённости реакция сильнее и асимметричнее; «ложноположительные» всплески чаще.
- **Через разливы (spillovers):** новости в одной компании/стране перекладываются на поставщиков, конкурентов, кред. спрэды и ETF.

Информационная среда **шумна и уязвима:** дубликаты, перепечатки, фрагментарные обновления, слухи, устаревшие материалы, риск дезинформации. Поэтому критичны **источник, время, подтверждения, версия новости** и видимая **цепочка цитирований**.

#### ▼ Почему это важно бизнесу

- **Время-к-инсайту:** оперативная приоритизация ленты и «почему важно» экономит часы аналитика и снижает упущенную выгоду.
- **Управление риском:** раннее выявление хвостовых событий, пересмотр лимитов, хеджирование до массового распространения.
- **Коммуникации и IR:** единая фактография с цитатами и таймлайном, пригодная для внутренних заметок и внешних комментариев.
- **Производительность контента:** черновики постов/заметок по актуальной повестке с проверяемыми ссылками и аккуратной атрибуцией.

Именно эту реальность — **поиск значимого в потоках, оценка его важности и сборка проверяемого контекста** — мы и предлагаем вам формально исследовать в рамках задачи RADAR.

## Образ решения

## MVP (решение)

Реализовать сервис «поиска и оценки горячих новостей», который на вход получает **временное окно** (например, последние 24 часа) и возвращает **топ-К событий** с оценкой «горячести» и черновиком публикации.

### ▼ Как может выглядеть выход

- `headline` — краткий заголовок;
  - `hotness ∈ [0,1]` — оценка «горячести»;
  - `why_now` — 1–2 фразы, почему важно сейчас (новизна, подтверждения, масштаб затронутых активов);
  - `entities` — компании/тикеры/страны/сектора;
  - `sources` — 3–5 проверяемых ссылок (оригинал, подтверждение, апдейт);
  - `timeline` — ключевые метки времени (первое сообщение → подтверждение → уточнение);
  - `draft` — черновик для поста/заметки (заголовок, лид-абзац, 3 буллета, цитата/сноска);
  - `dedup_group` — ID кластера дубликатов/перепечаток.
- 

## Питчинг (презентация)

**Цель:** показать, что сервис устойчиво находит **значимые** сюжеты, объясняет «почему сейчас» и отдаёт **готовый к использованию** черновик с проверяемыми ссылками.

**Регламент:** 5 минут презентация + 3 минуты Q&A.

### ▼ Рекомендуемая структура слайдов (под питч)

1. **Цель и критерии «горячести».** Одним слайдом: как вы формализовали «внезапность, подтверждения, масштаб, скорость».
2. **Источники и политика качества.** Откуда собираете, как боретесь с дубликатами/перепечатками и как оцениваете репутацию источников.

3. **Пайплайн.** Сбор → дедуп → ранжирование → контекст → черновик; где применяются модели, где правила, где кэш.
4. **Примеры топ-сюжетов.** 2–3 реальных кейса: заголовок, `hotness`, `why_now`, `timeline`, 3–5 ссылок; 1 «ложноположительный» с объяснением фильтра.
5. **Черновики.** Скрин/маркап: заголовок, лид, 3 буллета, цитата; как обеспечили проверяемость ссылками.
6. **Эффективность и риски.** Время ранжирования/черновика, кэш/эмбединги; как снижаете риск дезинформации и устаревших материалов.
7. **Выводы/next steps.** Что уже стабильно, что улучшать (например, мульти-язык, алертинг, богаче таймлайн).



## Оценка

### ▼ Питч

#### Оценивание.

Каждый жури ставит балл в диапазоне [0, 5] по каждому критерию. Оценки усредняются по жури и агрегируются по указанным ниже весам (сумма весов = 100%).

#### Критерии.

1. **Функциональность — 40%:** система стабильно закрывает весь цикл *сбор → дедупликация → ранжирование «горячести» → контекст → черновик* на реальном окне новостей, показывая топ-сюжеты и их проверяемые источники.
2. **Архитектура и пайплайн — 20%:** решение модульно и воспроизводимо (чёткие контуры, версии, повторяемые команды), устойчиво к шуму и легко расширяется новыми источниками/правилами.
3. **Уникальность/инновационность/вклад — 15%:** в решении есть оригинальная идея/приём, подтверждённый экспериментом и дающий измеримый эффект сверх бейзлайна.

4. **Достоверность и проверяемость — 15%:** каждый тезис «почему важно сейчас» привязан к первоисточнику/таймлайну и подтверждениям, с явной политикой борьбы с дезинформацией и перепечатками.
5. **Презентация/коммуникация — 10%:** логичная история в пределах регламента с читаемыми графиками и точными ответами без лишней риторики.