

## Лингвистика и цифровые технологии: участие в X Международном Невском форуме



28 июня 2026 года в залах Мариинского дворца — исторической резиденции Законодательного Собрания Санкт-Петербурга — состоялся **X Международный Невский форум**. В рамках форума прошел **круглый стол «Роль молодого ученого в мире ИИ-технологий»** с участием представителей советов молодых ученых ведущих университетов и научных институтов Северной столицы.

Площадка, принявшая дискуссию, говорила сама за себя: Мариинский дворец — место, где принимаются законодательные решения, определяющие жизнь города и страны. Именно здесь молодые исследователи из самых разных областей знания — математики, истории, экономики, лингвистики, государственного управления — обсуждали вопросы, от ответов на которые зависит будущее науки: как ИИ-технологии меняют конкретные исследовательские поля, заменит ли искусственный интеллект ученого в ближайшие десять лет и какой будет научная среда к 2035 году.



С докладом «Трансформация лингвистической науки в эпоху больших языковых моделей: от корпусных методов к ИИ-анализу» выступили к.ф.н., с.н.с. **Королькова Мария Денисовна** (Институт лингвистических исследований РАН) и к.ф.н., доц. **Дмитриев Александр Владиславович** (Гуманитарный институт СПбПУ Петра Великого / ИЛИ РАН). Центральной темой доклада стала реально работающая модель научного сотрудничества между академическим институтом РАН и университетом технического профиля — модель, воплощенная в конкретных магистерских и студенческих проектах.

Докладчики представили семь завершенных и анонсируемых проектов, выполненных студентами ВШЛиП СПбПУ в партнерстве с ИЛИ РАН:

— **«Не дай языку умереть»** — чат-бот для сохранения науканского языка: 5 634 словарных записи, 73 аудиозаписи носителей, архитектура, масштабируемая на другие малоресурсные агглютинативные языки.

— **«Классика в цифре»** — редизайн и веб-реализация корпуса Corpus Latino-Rossicum по техническому заданию ИЛИ РАН: три языка, четыре функциональных модуля, интерфейс, готовый к подключению производственного API института.

— **«Эфир под микроскопом»** — первый аннотированный корпус речевых ошибок профессиональной медиаречи: около 1 000 единиц, охват архива 2006–2025 годов, источник — рукописный архив ИЛИ РАН; морфологические ошибки составили 35,9% от общего числа.

— **«Мосты между языками»** — верифицированный параллельный китайско-русский корпус

для дообучения нейросетевых моделей машинного перевода в специализированных доменах.

— **«Умная книжная полка для малышей»** (KidLit) — первый сбалансированный корпус русскоязычной детской литературы 0+ (120 текстов, 7 языков-источников), NLP-пайплайн из пяти модулей, корреляция с экспертными оценками  $r \geq 0,7$ ; подана заявка в Роспатент на метрику возрастного соответствия.

— **«Живое слово в музейном каталоге»** (анонс) — совместный проект Лаборатории цифровой лингвистики СПбПУ и Отдела диалектной лексикографии и лингвогеографии ИЛИ РАН: автоматическое извлечение и лингвогеографический анализ диалектной лексики из описаний Государственного каталога Музейного фонда РФ (50 млн записей). Первый в мире проект на пересечении диалектного NLP и музейной информатики.



Доклад М.Д. Корольковой и А.В. Дмитриева показал: союз академической лингвистики РАН и инженерной школы политехнического университета — это не декларация, а работающий механизм подготовки исследователей нового типа. Студенты решают реальные задачи институтов РАН, институты получают воспроизводимые инструменты и верифицированные ресурсы. Результат — не учебные упражнения, а научные продукты, готовые к применению в исследовательской практике.

X Международный Невский форум подтвердил: разговор о будущем науки в эпоху ИИ наиболее продуктивен тогда, когда за столом сидят люди, уже строящие это будущее своими руками.