ЦРТ | ГРУППА КОМПАНИЙ

MOSCOW FORENSICS DAY '25
12.09.25

From audio and video data to evidence

AI analytics, biometrics, spoofing and deepfake detection

Speaker: Mikhail Inkin
Head of project group

16+

About us

ЦРТ | ГРУППА КОМПАНИЙ

The STC group of companies develops products and solutions based on intelligent speech technologies, generative artificial intelligence, and face and voice biometrics



30+
years in the AI market

5000+
projects worldwide, including nationwide projects

One of the largest players in the AI market in Russia

Winner of international scientific competitions in speech recognition and biometrics

NIST CHiME CHALLENGE ASVspoof

750+
employees on staff

70+
R&D teams

20
PhDs and doctors of science

Talent pool

ИТМО

Корпоративная лаборатория Технологии человеко-машинного взаимодействия

www.speechpro.ru

LLM solutions from Sber

ЦРТ | ГРУППА КОМПАНИЙ

GIGA CHAT

Business process transformation

- Knowledge base search
- Margin income calculation
- Predictive equipment diagnostics
- Procurement assistant

GIGAIDE GIGA CODE GIT VERSE

www.speechpro.ru

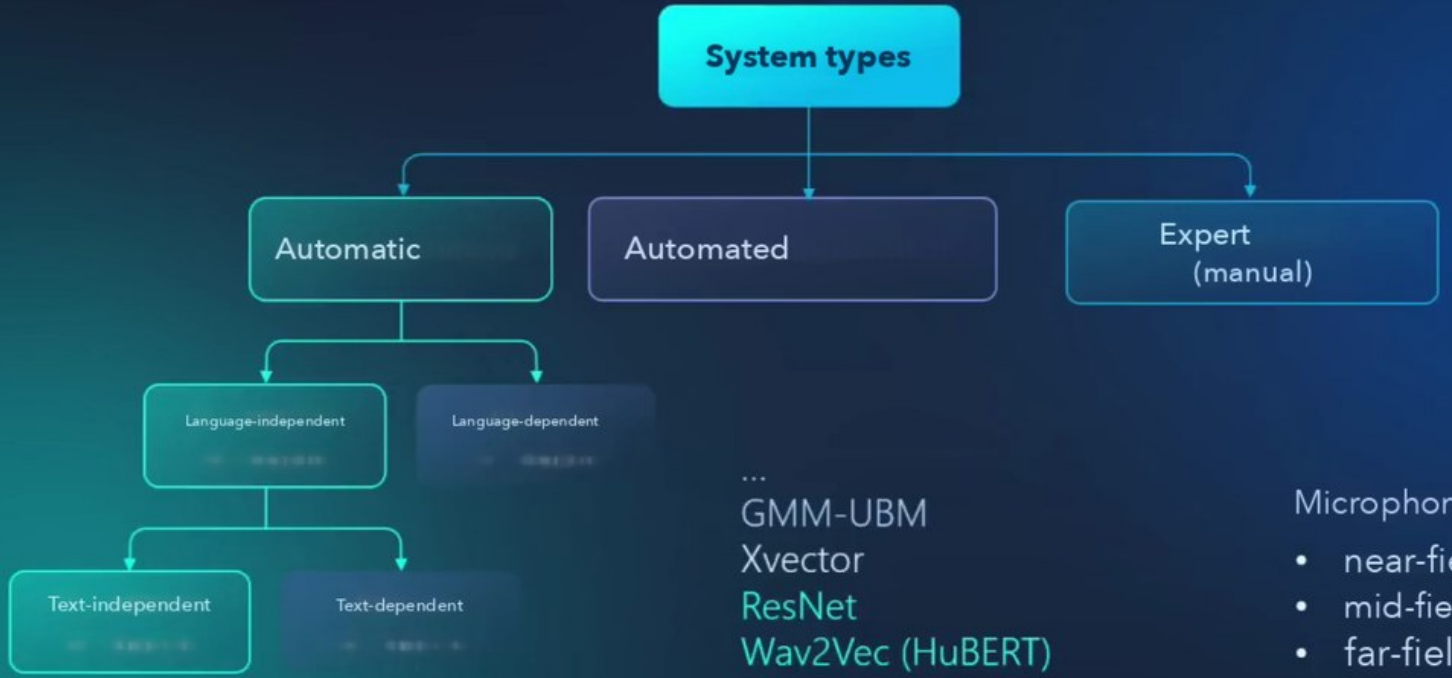
Ever-growing volumes of unstructured data



- Low efficiency of data search and processing
- High cost of scaling data processing systems
- Low speed of analysis and decision-making
- Loss of important and valuable information

www.speechpro.ru

Types of biometric systems



...
GMM-UBM
Xvector
ResNet
Wav2Vec (HuBERT)
...

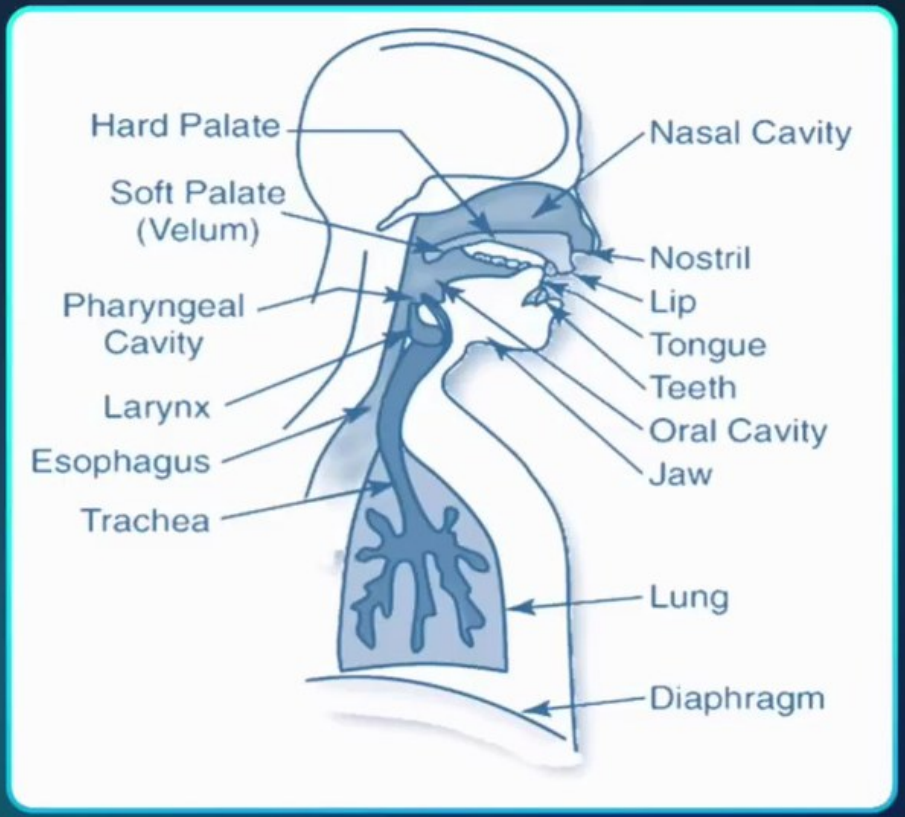
Microphone type:

- near-field
- mid-field
- far-field

www.speechpro.ru

Human speech synthesis

- Every person's voice is unique!
- More than 70 body parts are involved in producing the voice
- Not what we say matters, but who speaks
- More than 30 years of use in forensic examination

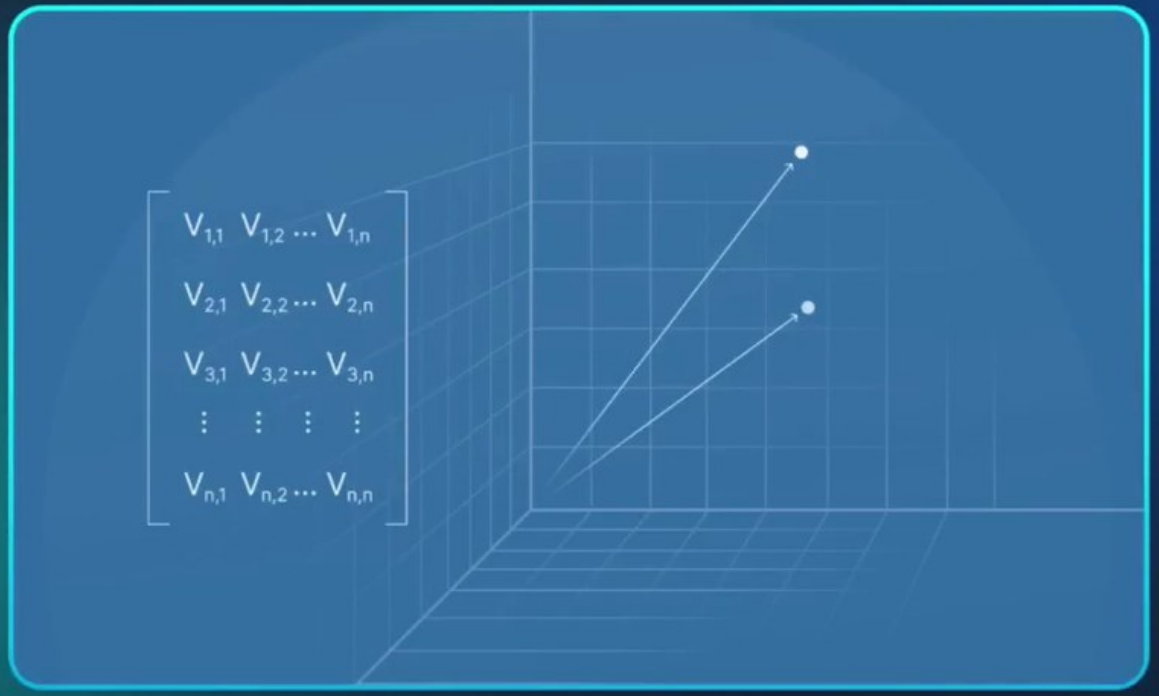


www.speechpro.ru

Speaker model

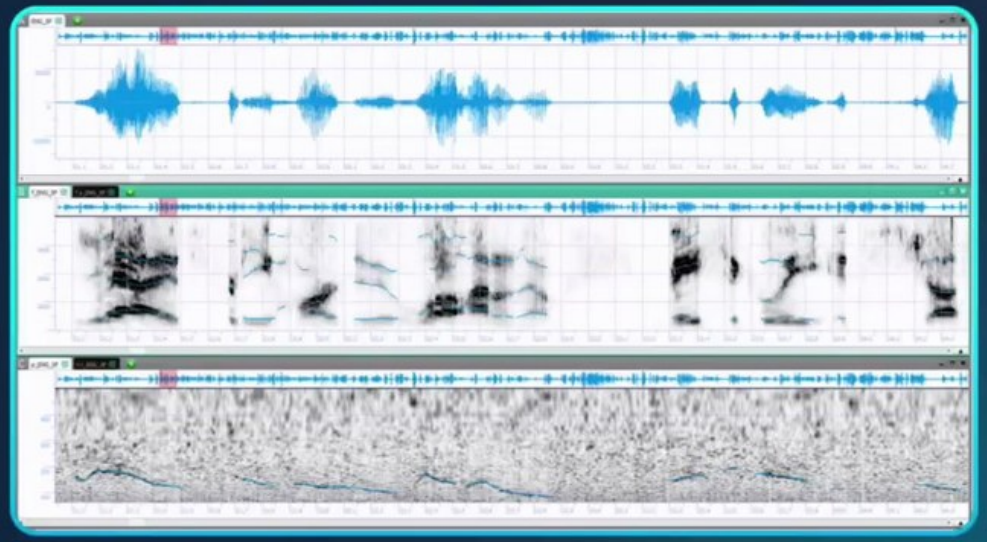
ЦРТ | ГРУППА КОМПАНИЙ

Feature vector, embedding



www.speechpro.ru

Identifying characteristics of the voice



Fundamental frequency (F0) (pitch, fundamental tone) – the rate of vocal fold vibration, is determined by the vocal tract characteristics and lies in the ranges (60-180 Hz M, 160-350 Hz F, and 200-650 Hz for children).

Formants – frequency maxima arising from the resonance of the human vocal tract.

Usually three formants (F1, F2, F3) are enough both to distinguish sounds and to recognise the speaker.

Every formant is a function of all parts of the vocal tract, though the degree of their influence is not the same in each case.

Challenges of the task

ЦРТ | ГРУППА КОМПАНИЙ

Causes of data variability



Speaker-related

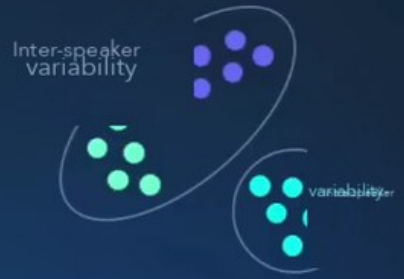


Technological

Reading, spontaneous speech, monologue, dialogue, group conversation

...

Communicative



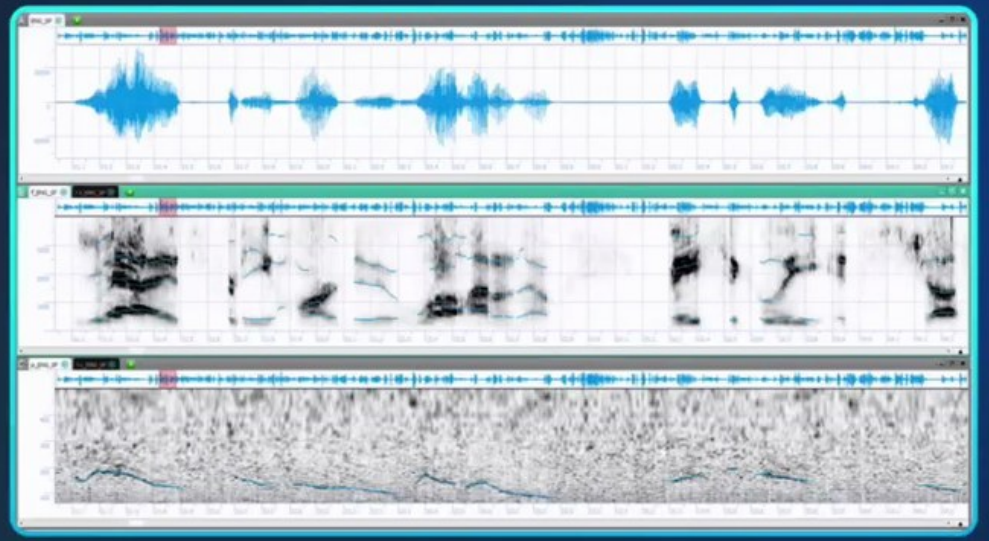


[1] - Speaker recognition by machines and humans [Hansen J.H.L., Hasan T., 2015]

www.speechpro.ru

Approaches to speaker comparison

- Auditory
- Acoustic-phonetic
- Spectrographic:
  comparison of formants, F0, F0 harmonics
- Automatic

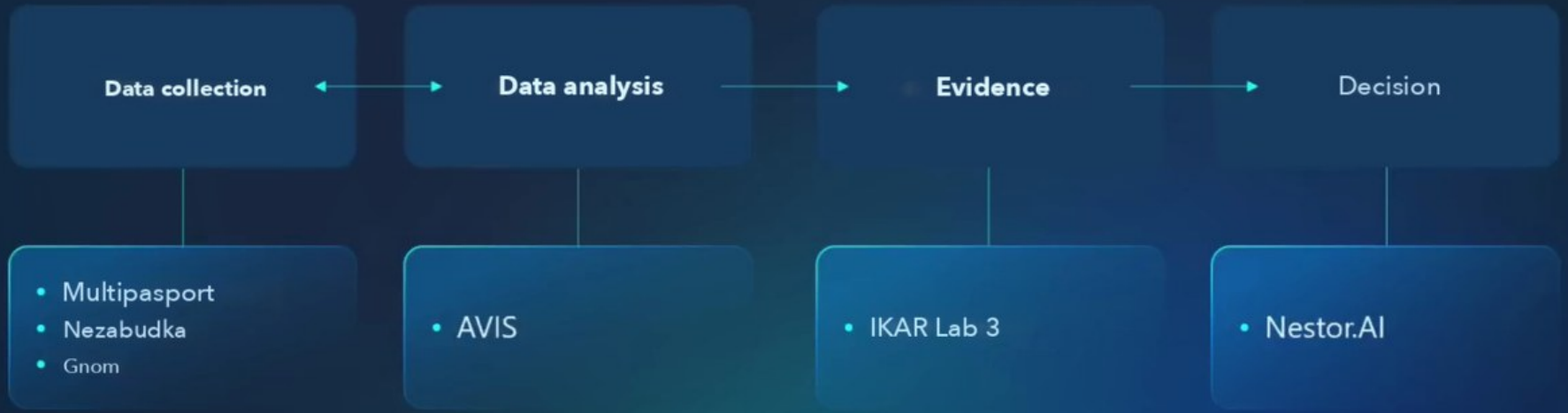


www.speechpro.ru

Solving a practical case



Product line and the stages of the forensic process



www.speechpro.ru

AVIS

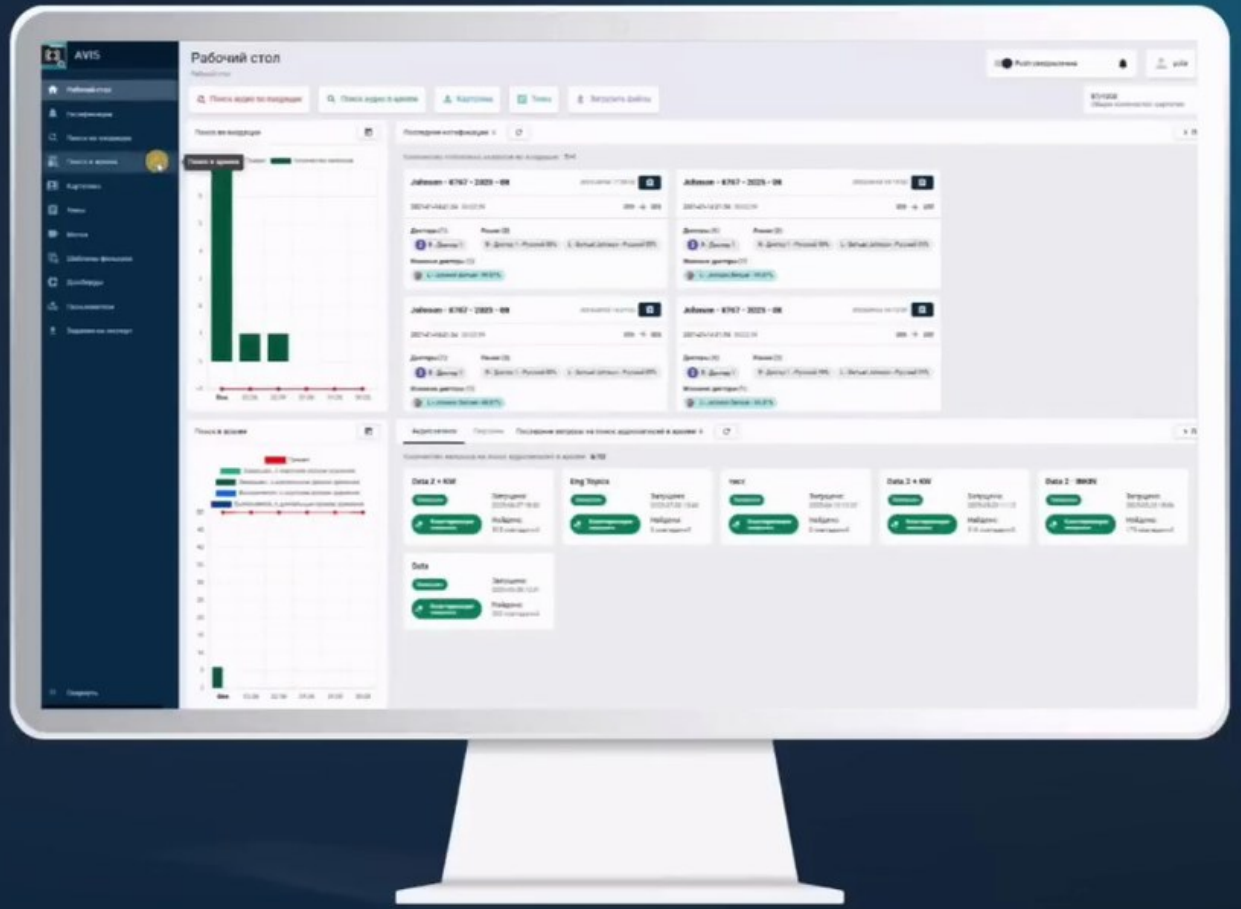
Products for public safety

ЦРТ | ГРУППА КОМПАНИЙ



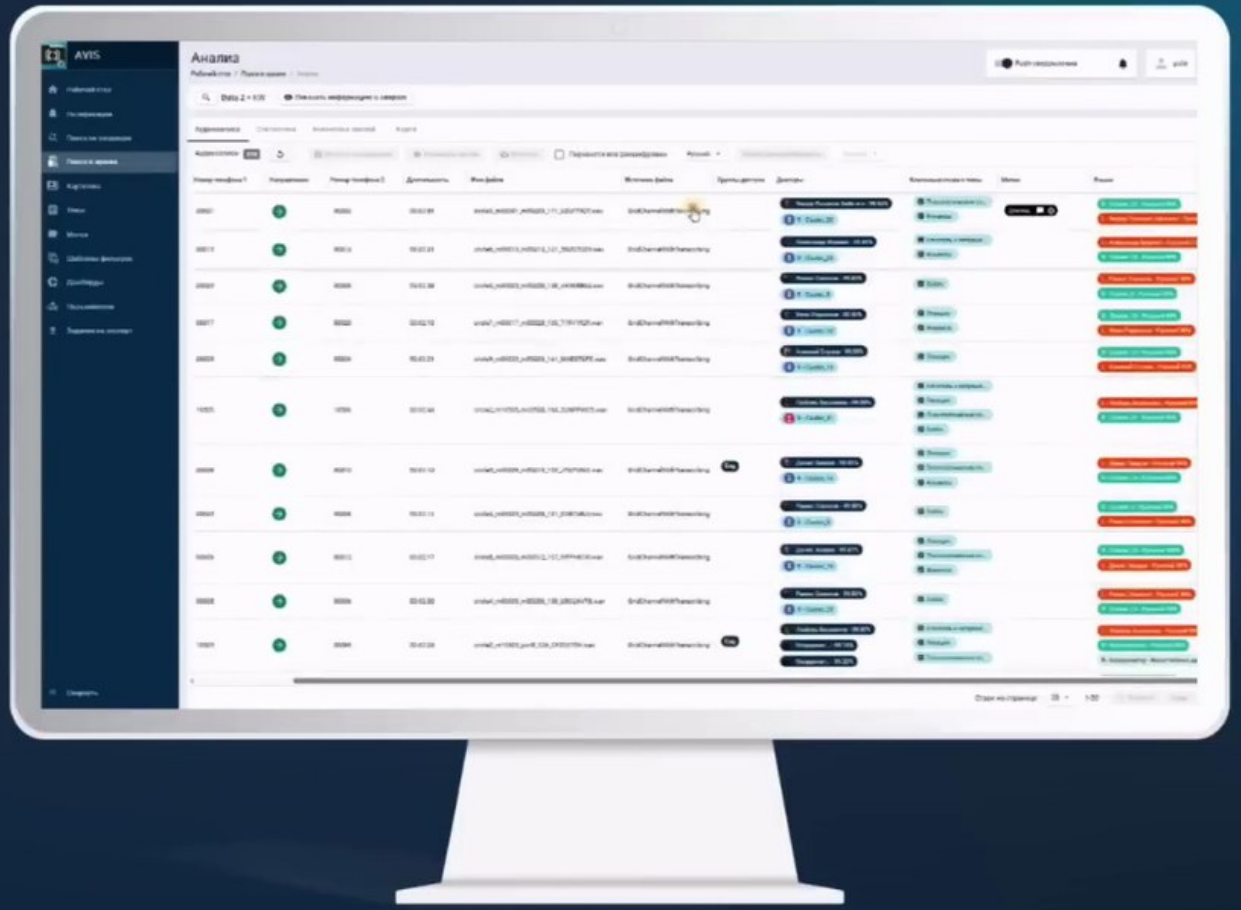
www.sc...tpro.ru

What happens to the data in AVIS



www.speechpro.ru

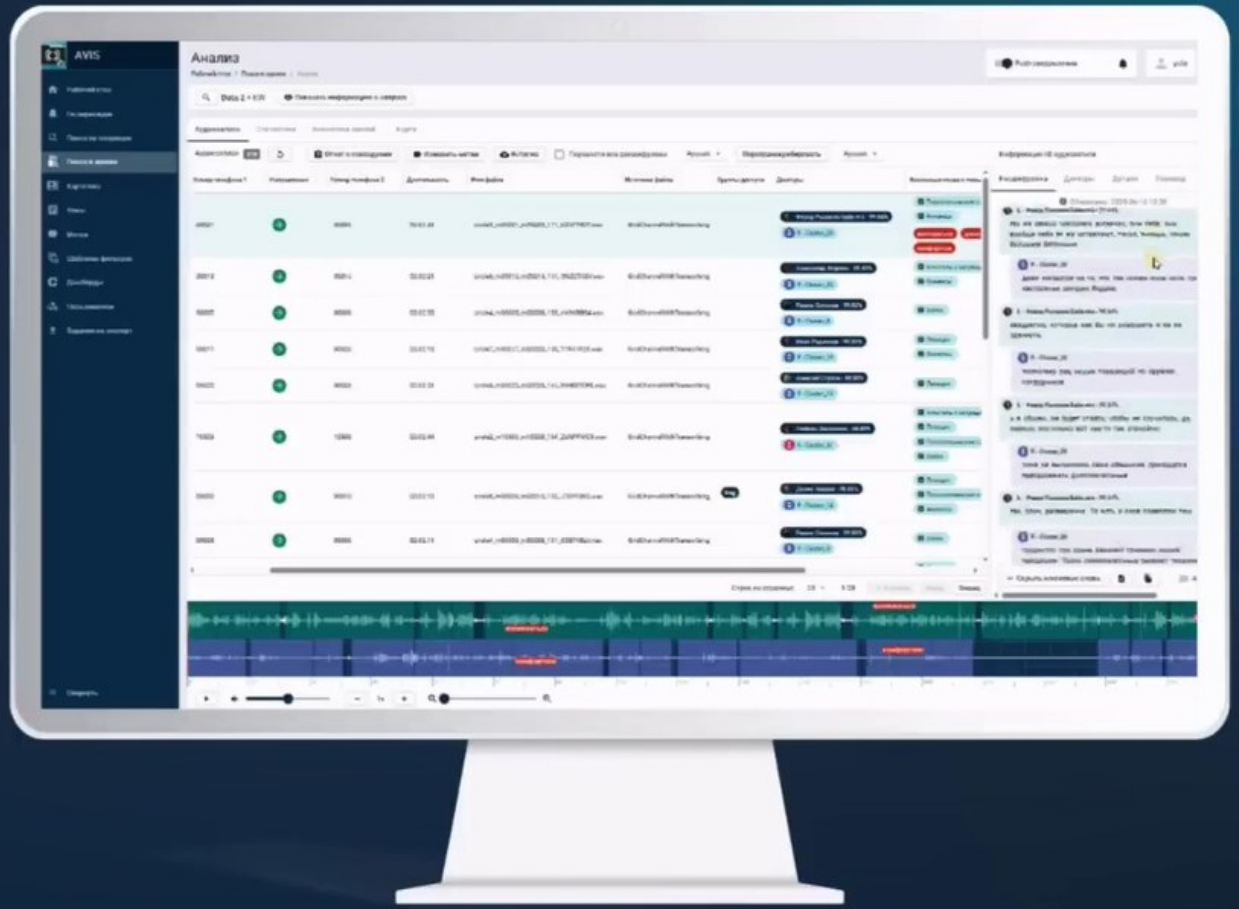
What happens to the data in AVIS



www.speechpro.ru

What happens to the data in AVIS

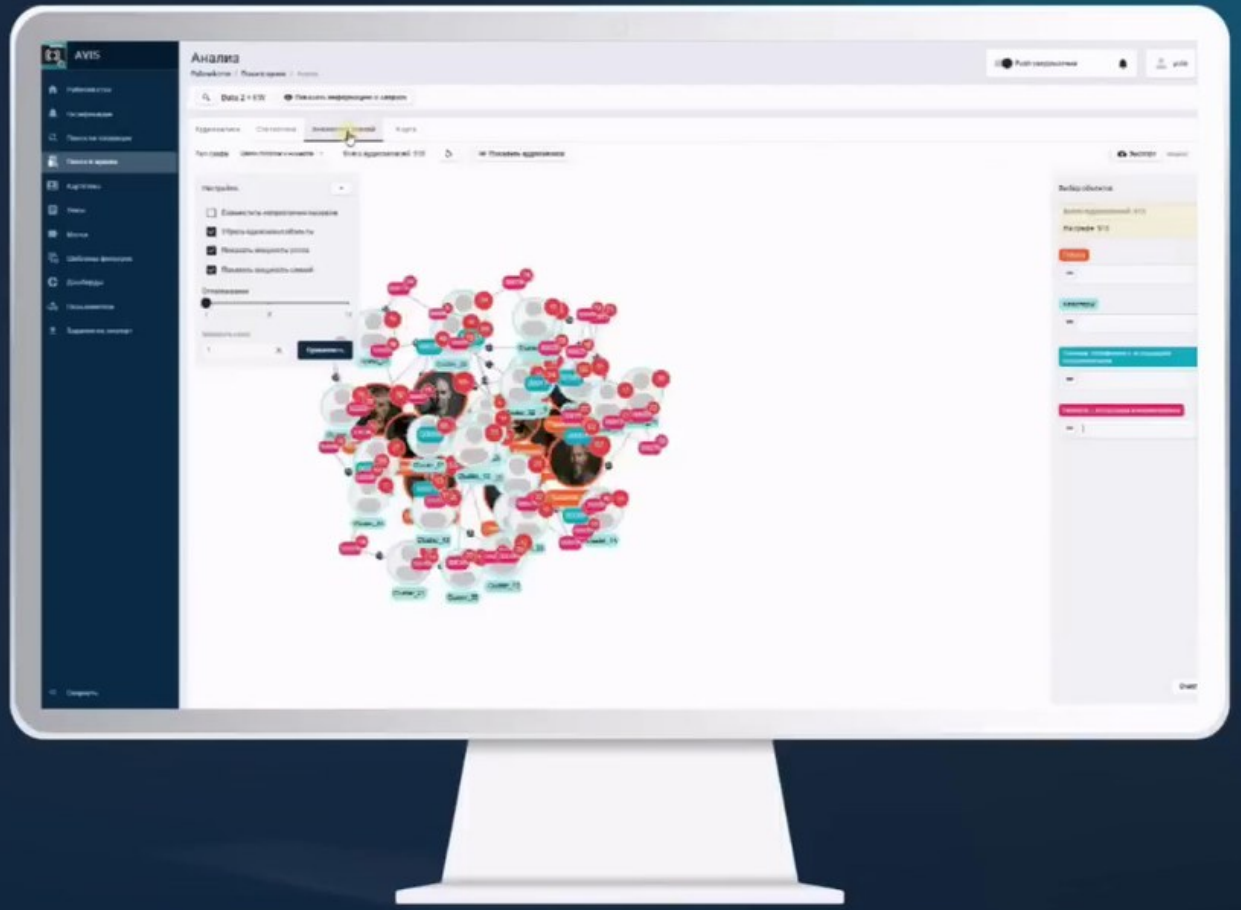
ЦРТ | ГРУППА КОМПАНИЙ



www.speechpro.ru

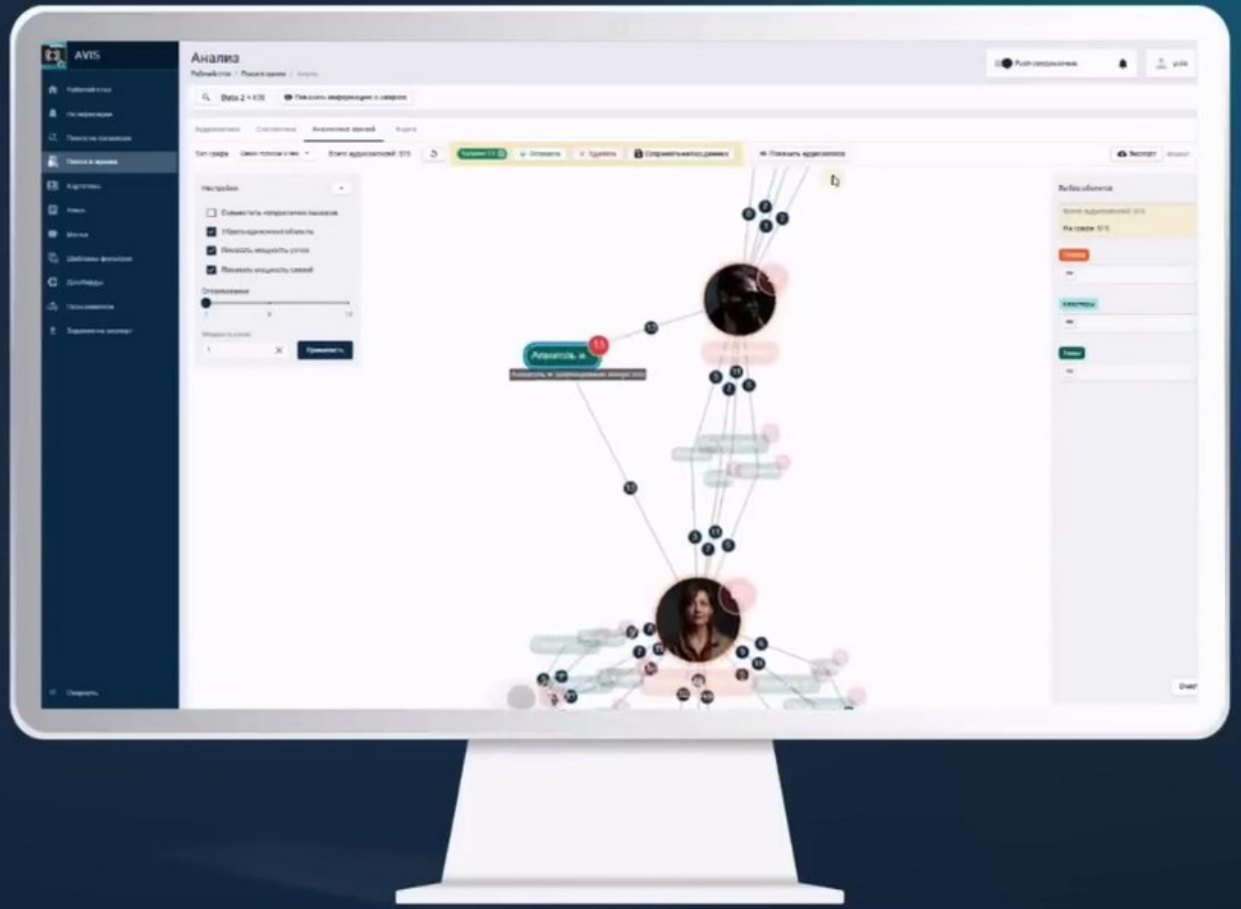
What happens to the data in AVIS

ЦРТ | ГРУППА КОМПАНИЙ



www.speechpro.ru

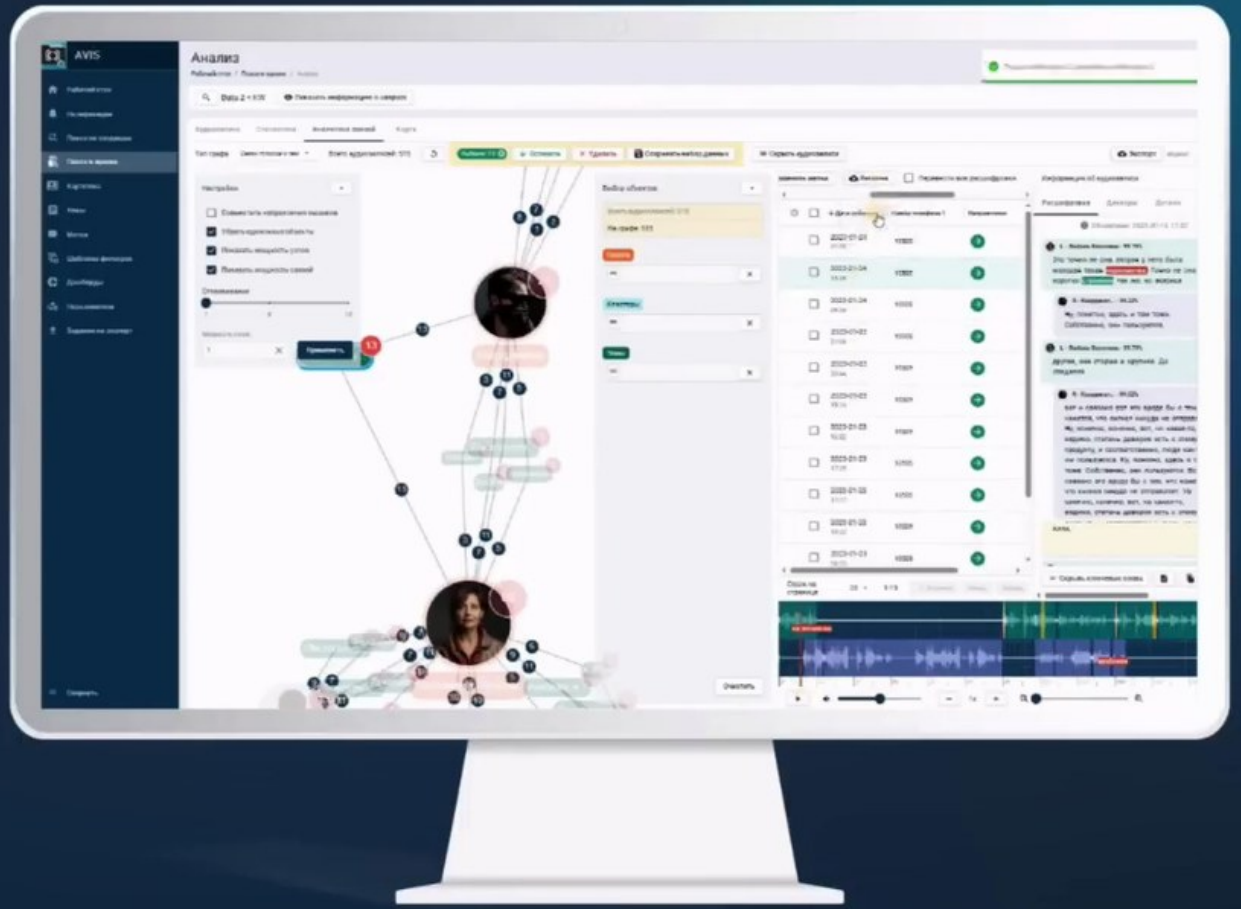
What happens to the data in AVIS



www.speechpro.ru

What happens to the data in AVIS

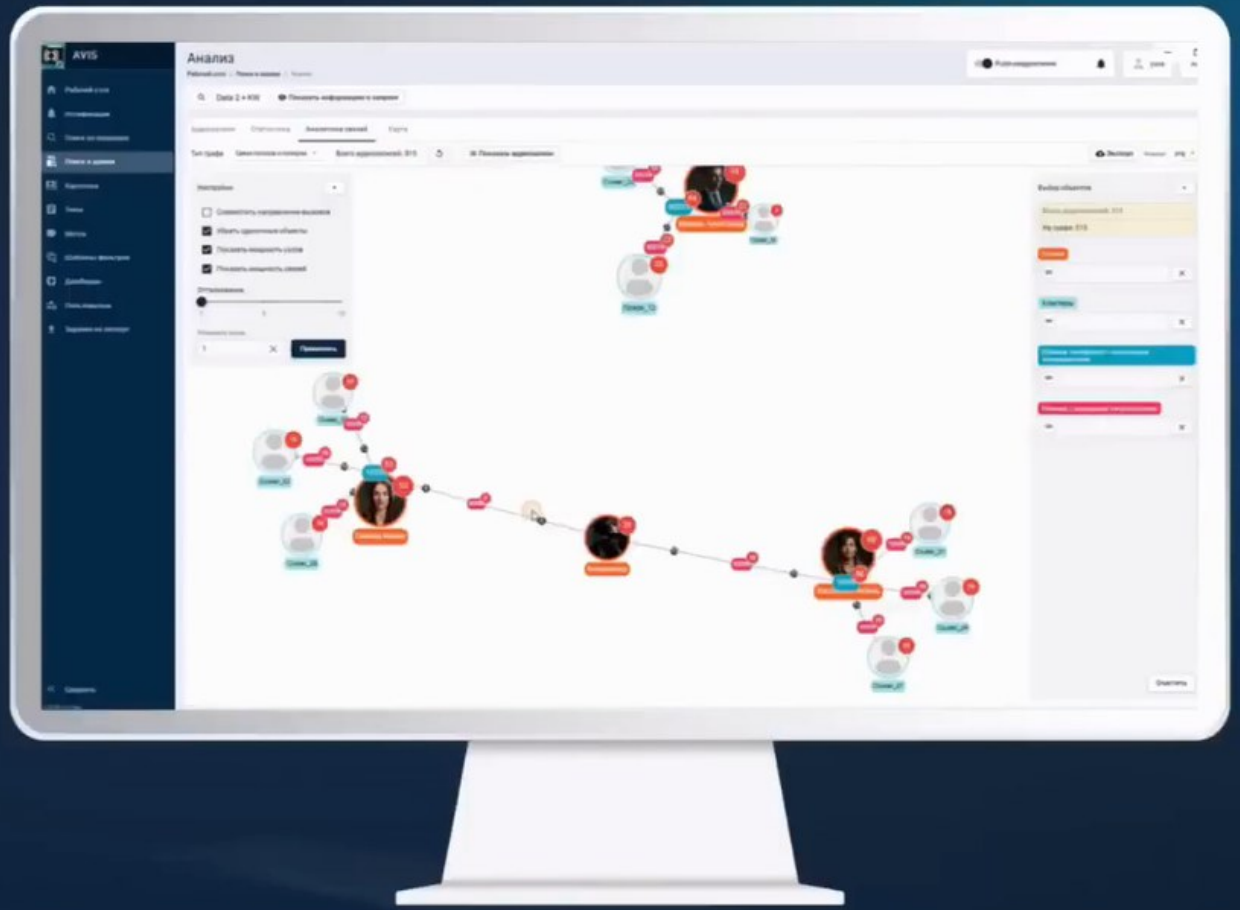
ЦРТ | ГРУППА КОМПАНИЙ



www.speechpro.ru

Biometric link graph

ЦРТ | ГРУППА КОМПАНИЙ



www.speechpro.ru

LLM in solving forensic tasks

ЦРТ | ГРУППА КОМПАНИЙ

Text generation from instructions

Creating coherent and grammatically correct text on a given topic

Question answering

Providing answers to questions based on information known to the LLM and on context

Text classification

Determining categories or labels for a given text (for example, sentiment analysis)

Personalisation of answers

Adapting answers and behaviour to the specific user

Information extraction

Picking out key facts and data from text

Conversation support

Holding dialogues with users, understanding and contextualising previous turns

Code generation and debugging

Writing code and fixing errors in it in various programming languages from a task description

Text summarisation

Condensed rendering of long texts, picking out the main ideas

Text translation

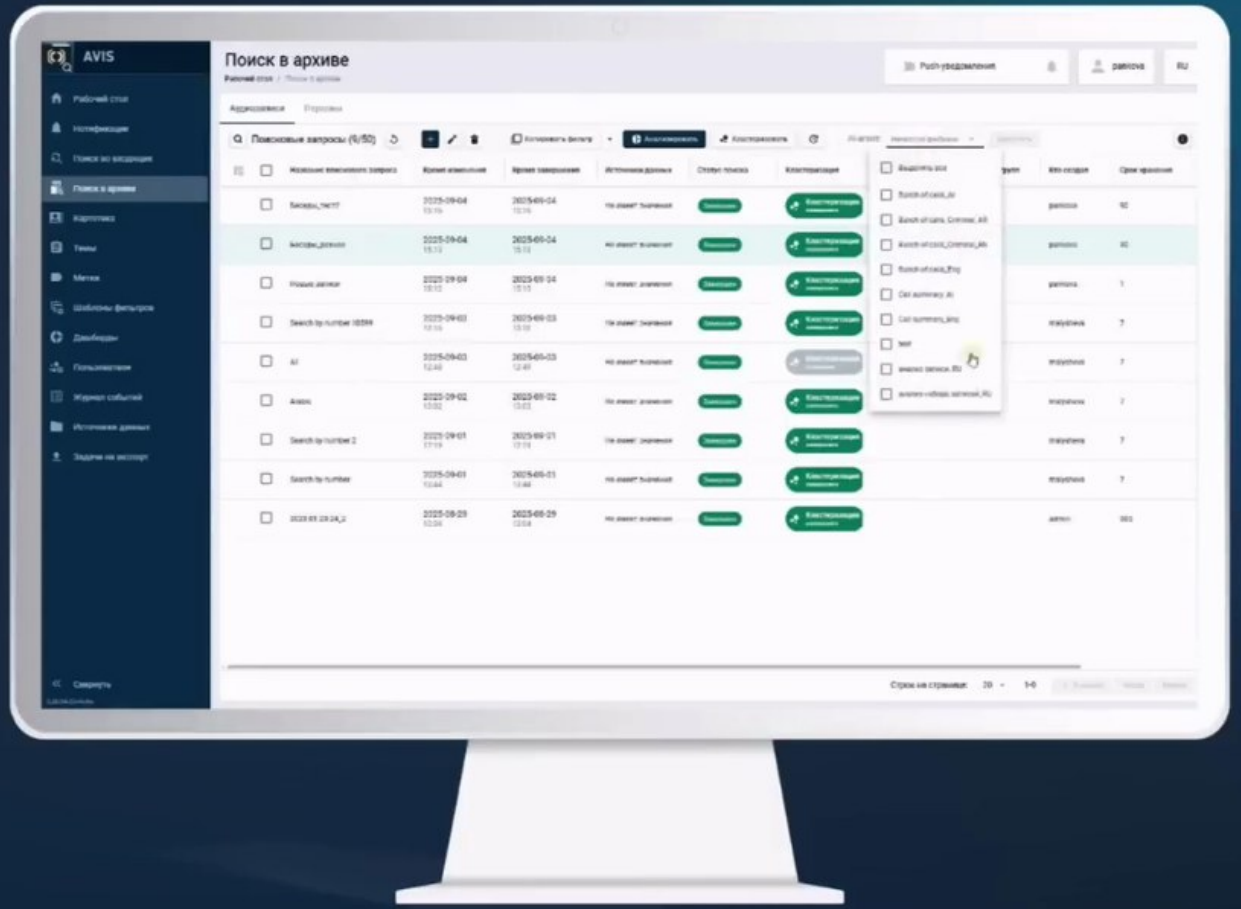
Translating text from one language into another

Data interpretation

Explaining data analysis results, data visualisation

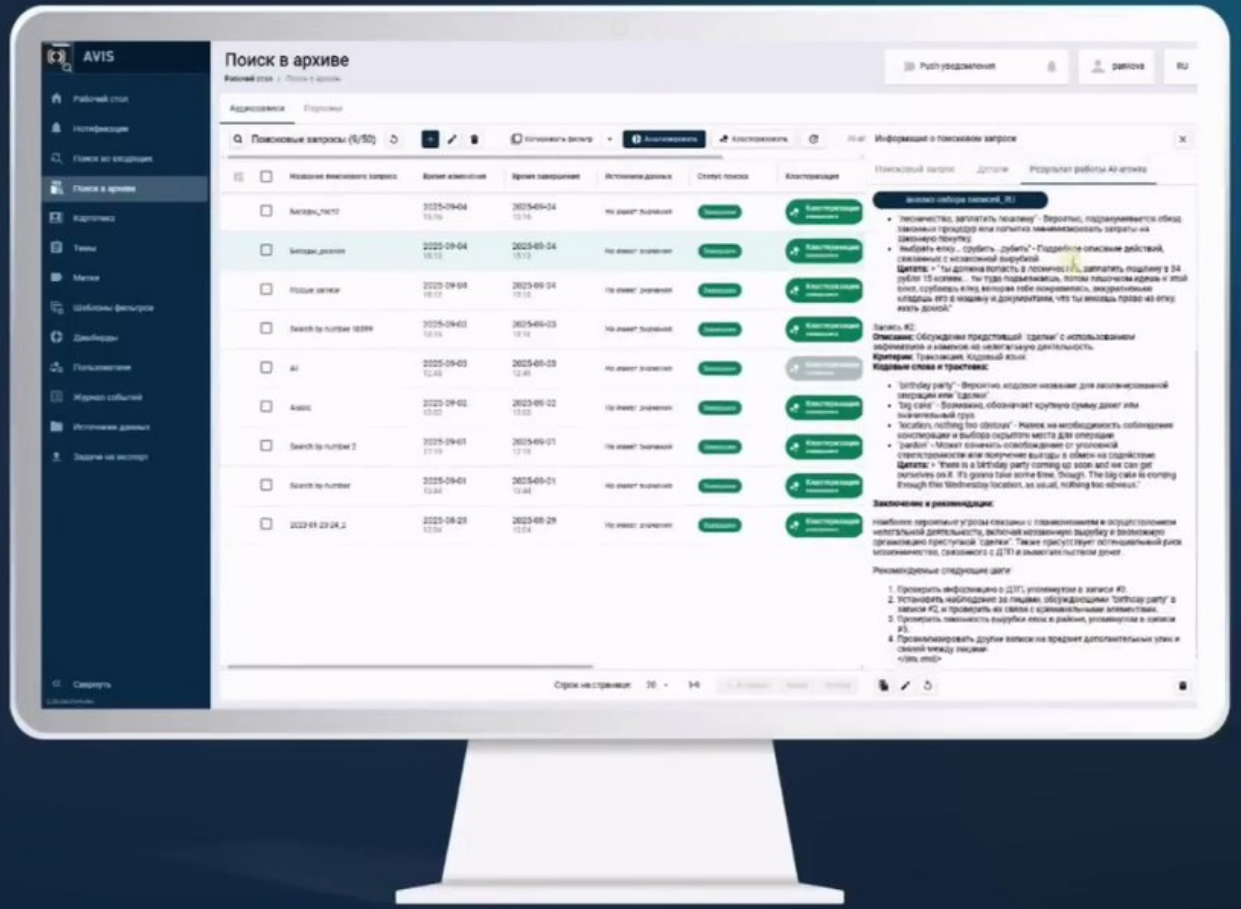
www.speechpro.ru

Bulk information processing



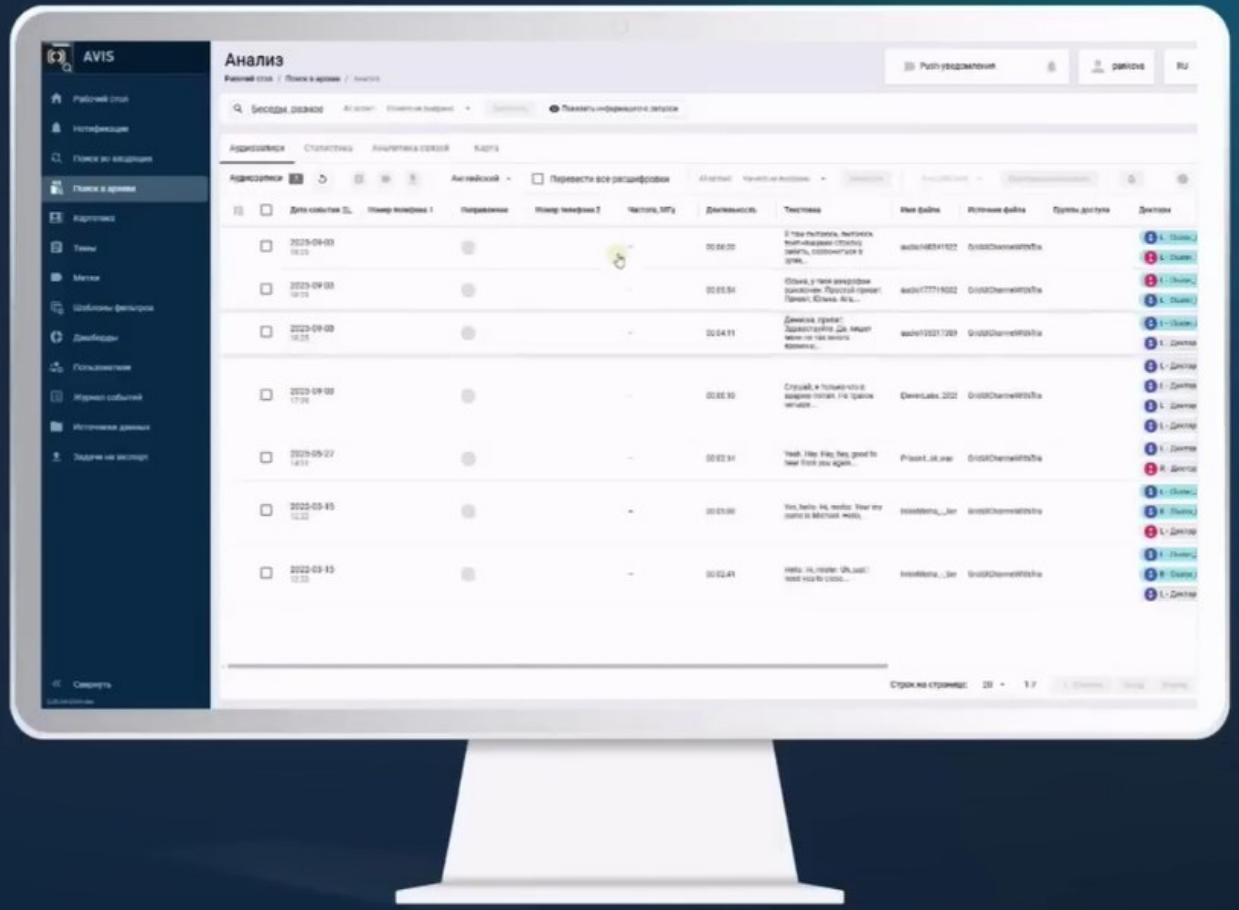
www.speechpro.ru

Bulk information processing



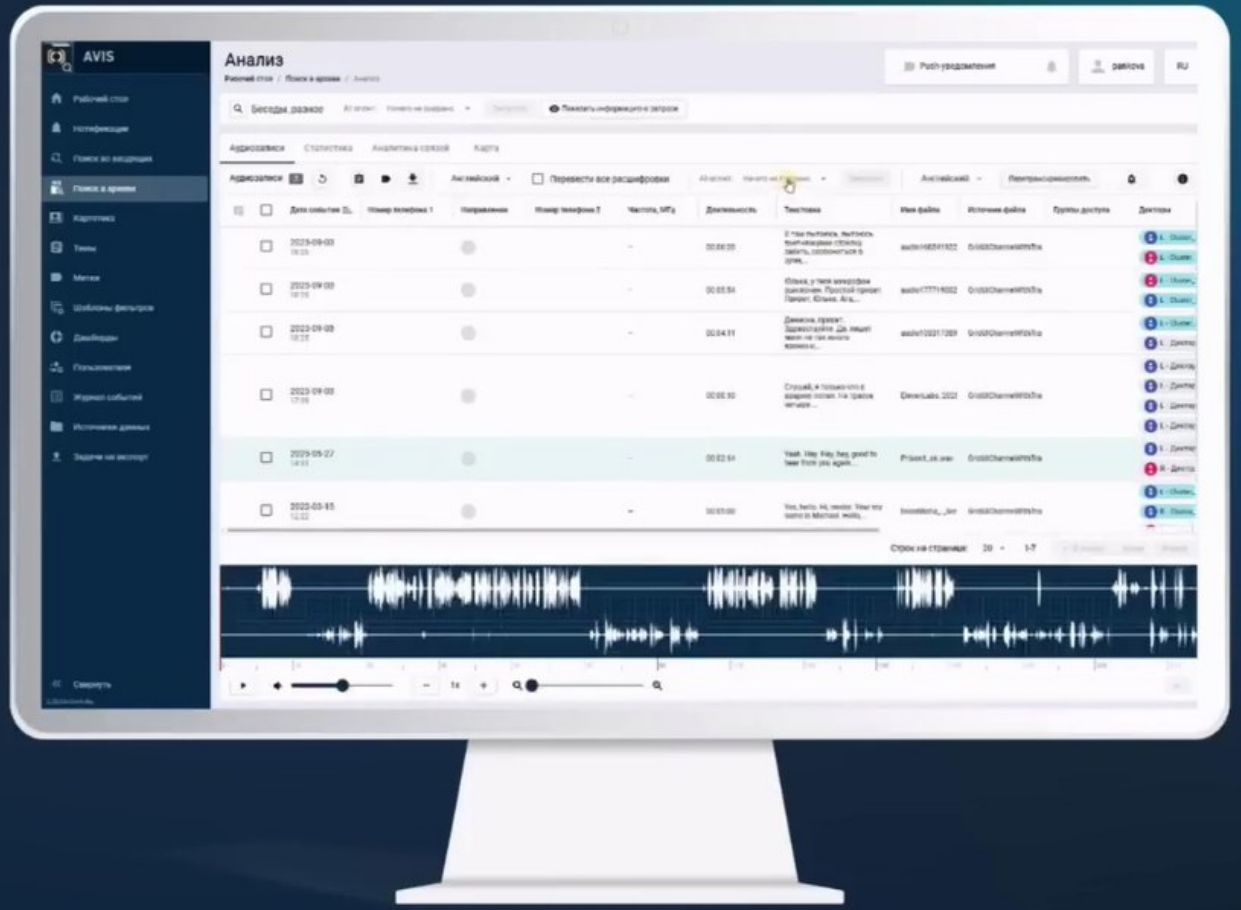
Bulk information processing

ЦРТ | ГРУППА КОМПАНИЙ



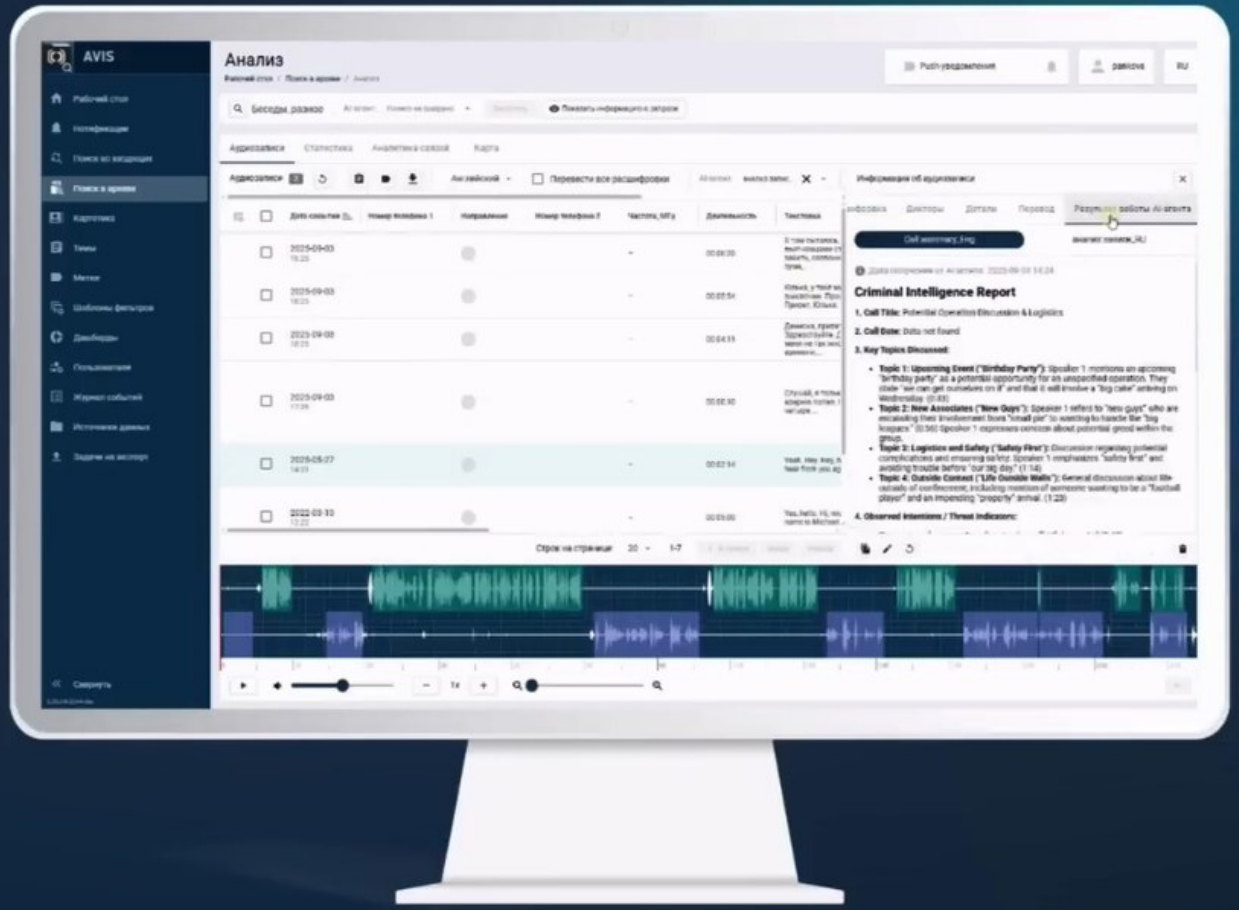
www.speechpro.ru

Bulk information processing



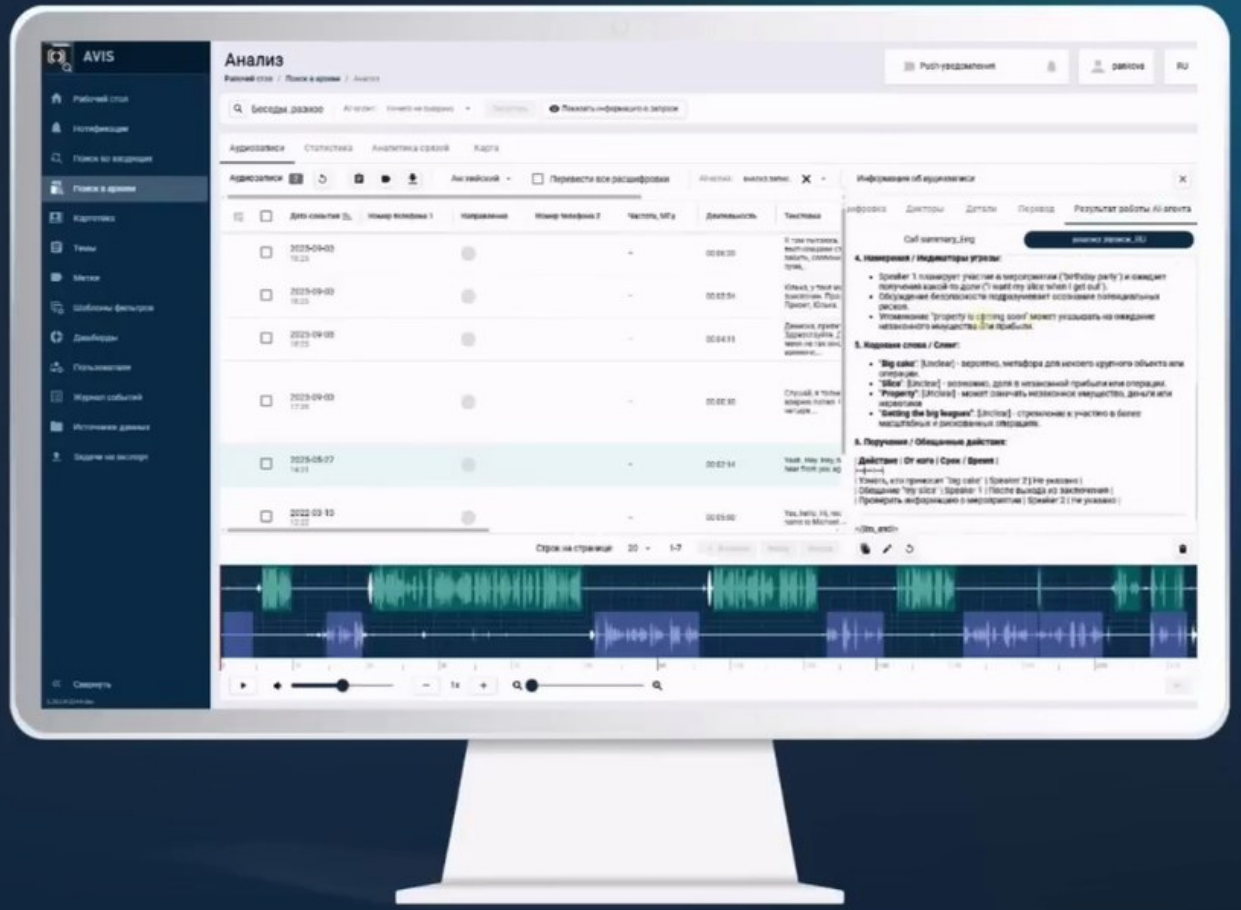
www.speechpro.ru

Bulk information processing



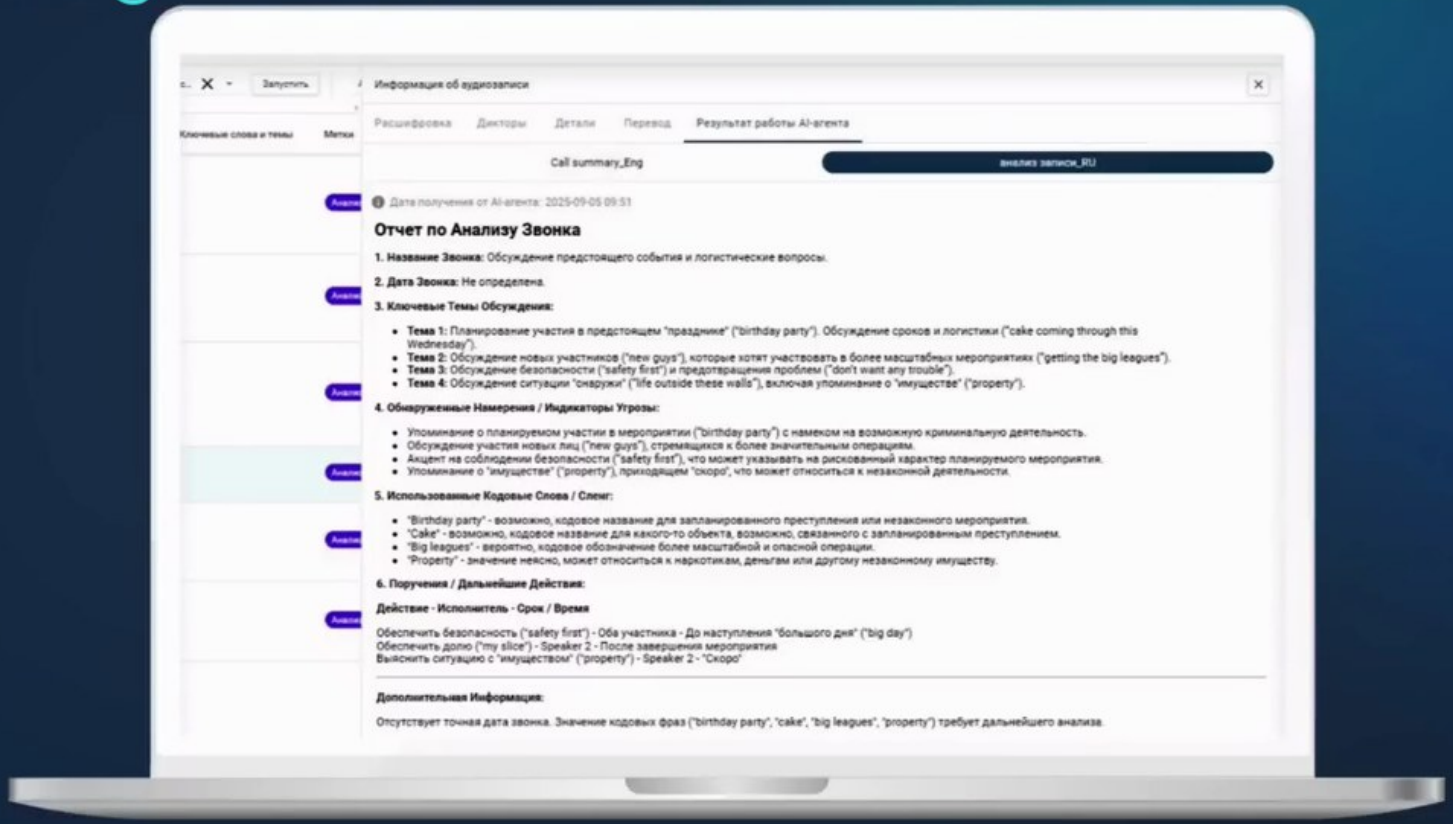
www.speechpro.ru

Bulk information processing



www.speechpro.ru

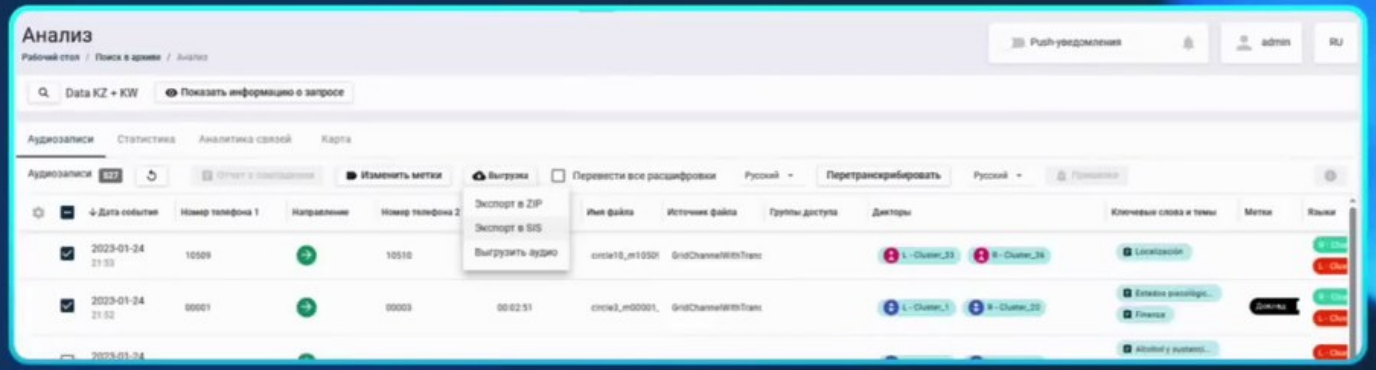
Auto-annotation of individual recordings



www.speechpro.ru

ЦРТ | ГРУППА КОМПАНИЙ

Exporting recordings to SIS



Preliminary analysis



www.speechpro.ru

Feature extraction and identification

- Automatic comparison algorithms
- Speech and noise detection
- Editing and processing of recordings
- Identification wizard
- Expert comparison methods
- Signal characteristics calculation
- Audio visualisation and analysis
- EdiTracker module
- Authenticity diagnostics
- Expert report
- "Super Loupe" module



ЦРТ | ГРУППА КОМПАНИЙ

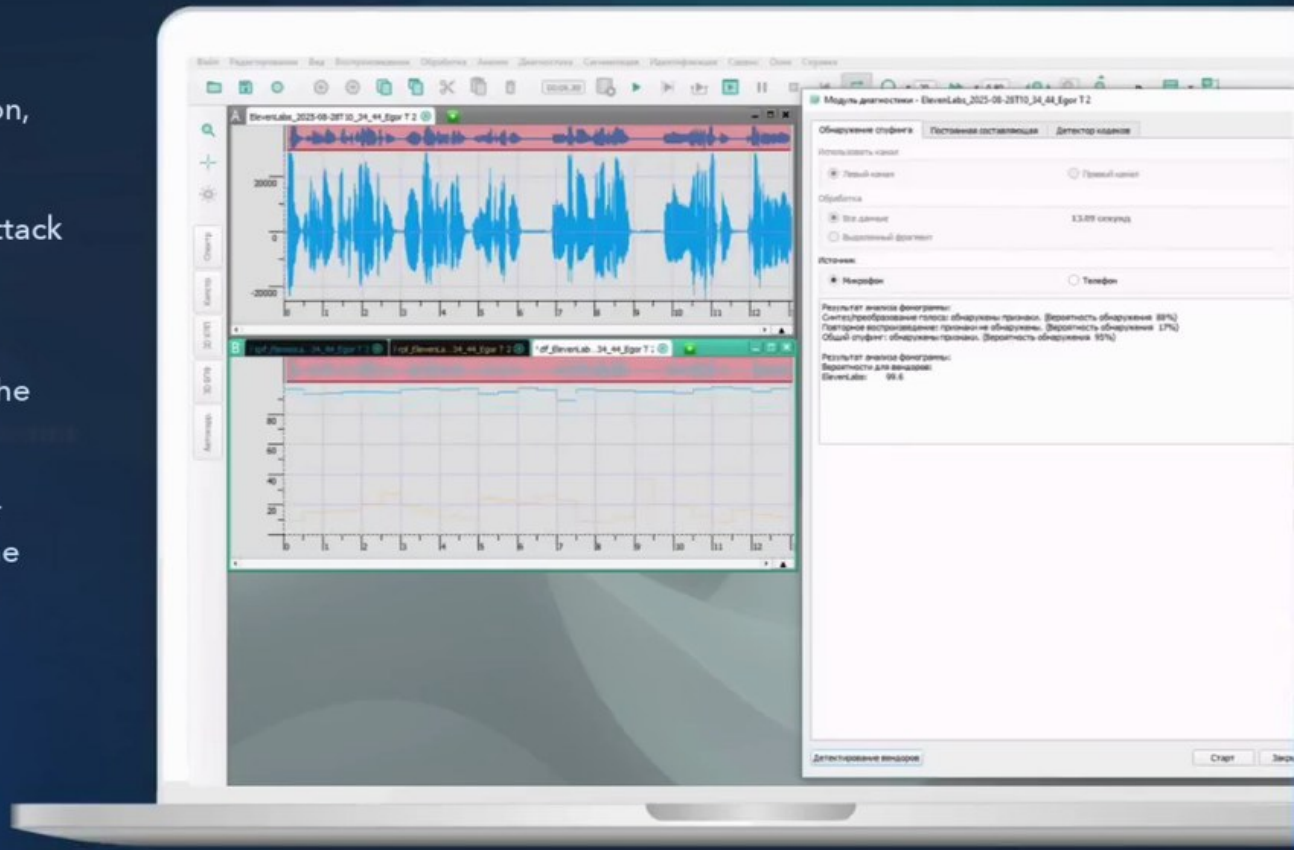
Example 2



Spoofing detector

ЦРТ | ГРУППА КОМПАНИЙ

- The detection module* estimates the probability of spoofing (synthesis/conversion, replay, overall score) in the recording.
- An overall score and separate scores by attack type are calculated.
- Three spoofogram graphs are produced, showing the sliding score values against the signal time.
- If the synthesis probability score is 50% or above, the extra function for identifying the synthesis vendor can be enabled.



* * Supplied under a separate licence

Spoofing check or authenticity check

ЦРТ | ГРУППА КОМПАНИЙ

A spoofing check – identifying and detecting forged media files with algorithms that analyse inconsistencies and with deepfake detection technologies.

An authenticity check – confirming the originality and integrity of digital content:

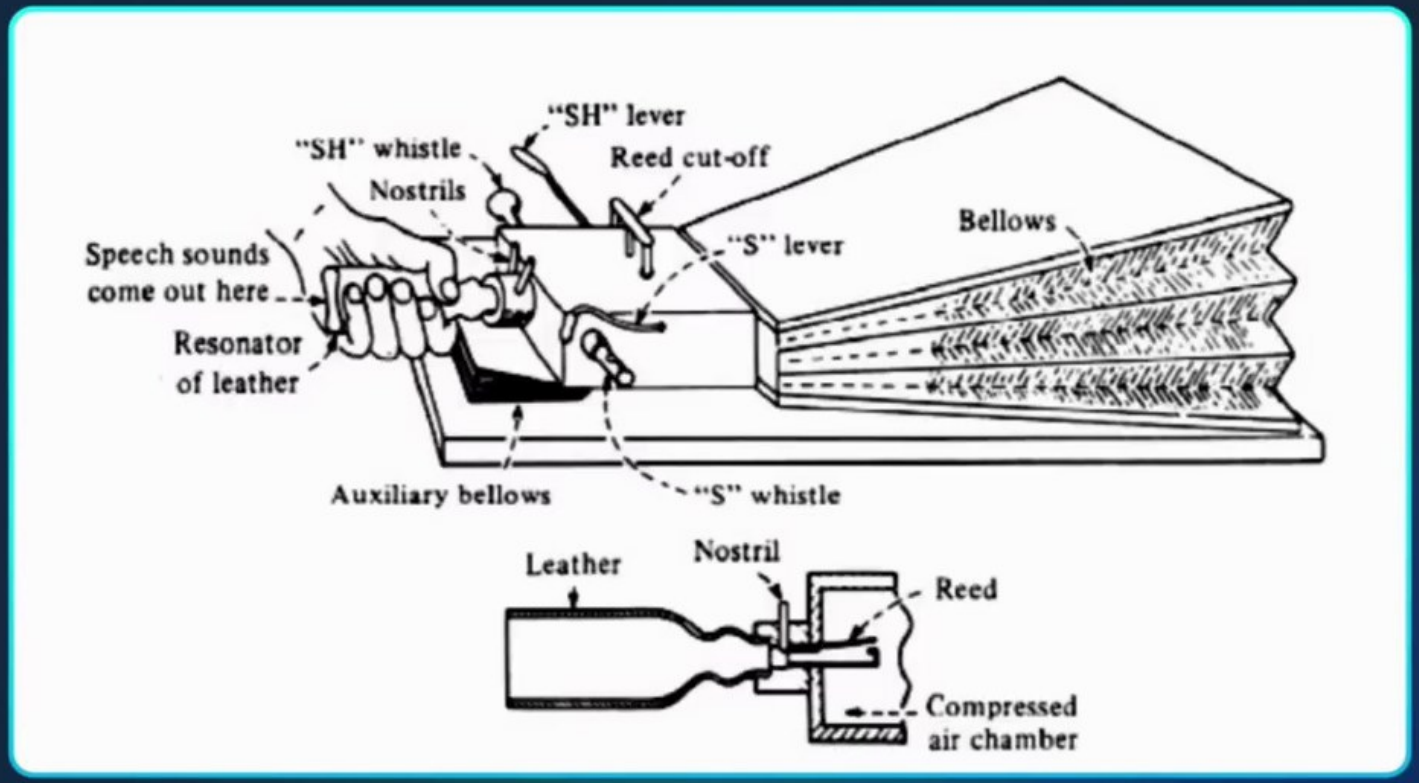
- metadata analysis;
- checking digital signatures and watermarks;
- comparison with trusted sources;
- deepfake detection.

Deepfake detection aims to find forged media files, whereas an authenticity check confirms the originality and integrity of digital content.



www.speechpro.ru

History of spoofing



Mark J. Jones, Rachael-Anne Knight (Eds.) The Bloomsbury Companion to Phonetics. 2013

Some examples

1920

1979

...

2000

2023

2024



Fidelity Voice Chess Challenger (1979) the first talking chess computer

https://www.ismenio.com/chess_fidelity_vcc.html

A simple speech generator

https://ieeexplore.ieee.org/document/861820

Simple AI voice cloning

Google search: voice clone



Advanced voice cloning systems with emotions

https://www.cnn.com/2024/01/22/politics/fake-joe-biden-robocall/index.html

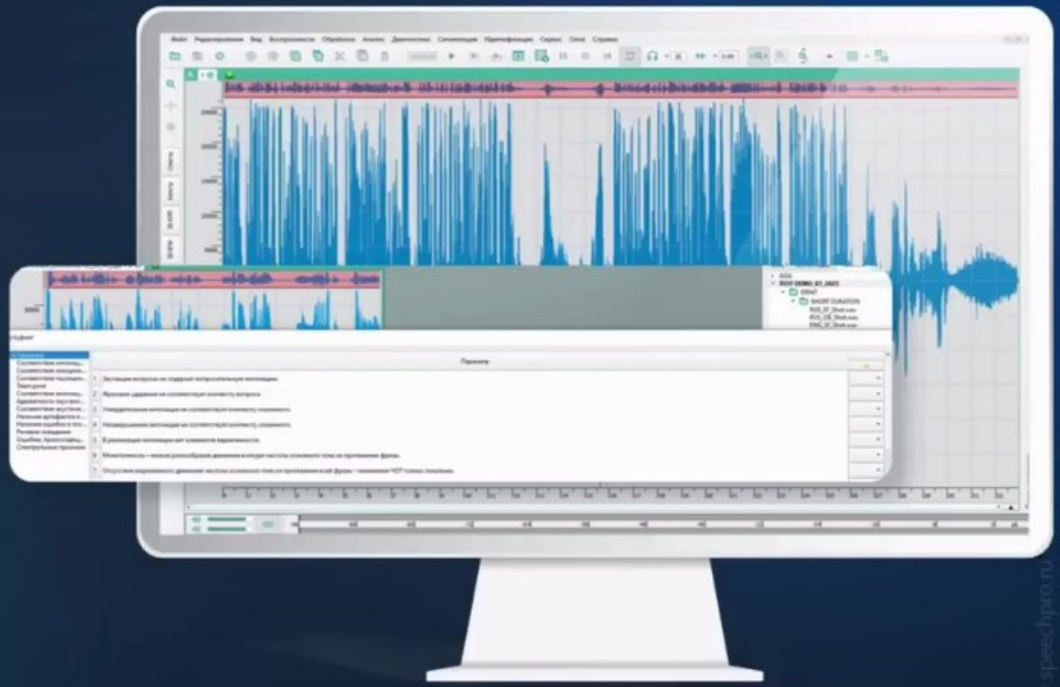
SIS capabilities

Expert method for detecting spoofing

Spoofing assessment

The automatic detector produces an initial score for the recording

- The expert records spoofing features in an interactive artefact table*
- The result – a report saved into the project



* * Supplied under a separate licence

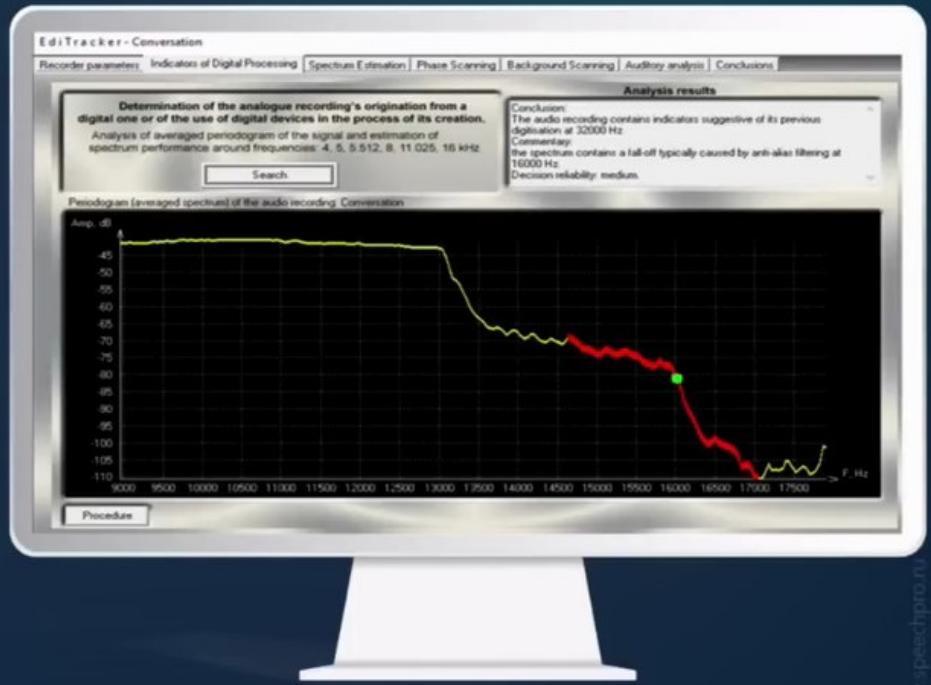
www.speechpro.ru

SIS capabilities

EdiTracker module

EdiTracker makes it possible to find signs of a breach of authenticity using the following methods:

- Examining the technical characteristics of the recording device
- Detecting traces of digital signal pre-processing
- Detecting stationary harmonics and scanning their phase
- Background noise scanning
- Auditory analysis



www.speechpro.ru

ЦРТ | ГРУППА КОМПАНИЙ

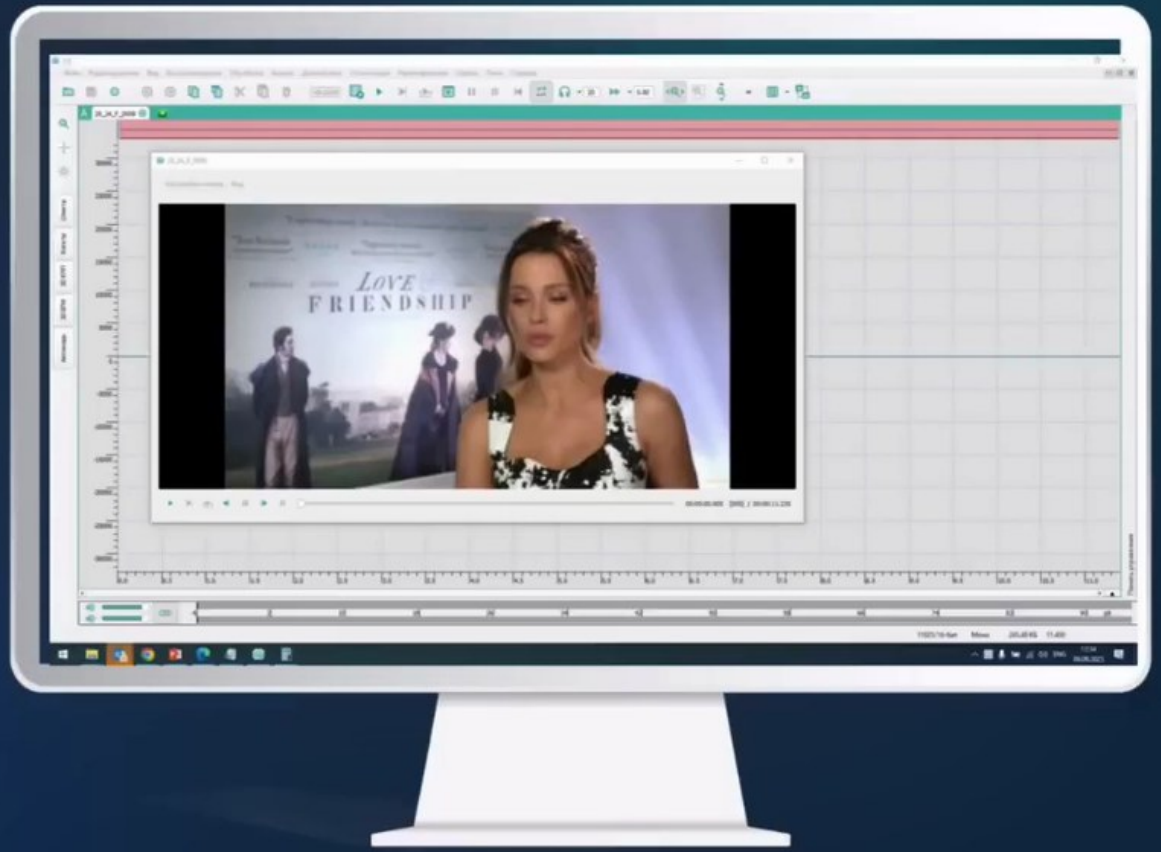
Example 3





Deepfake analysis

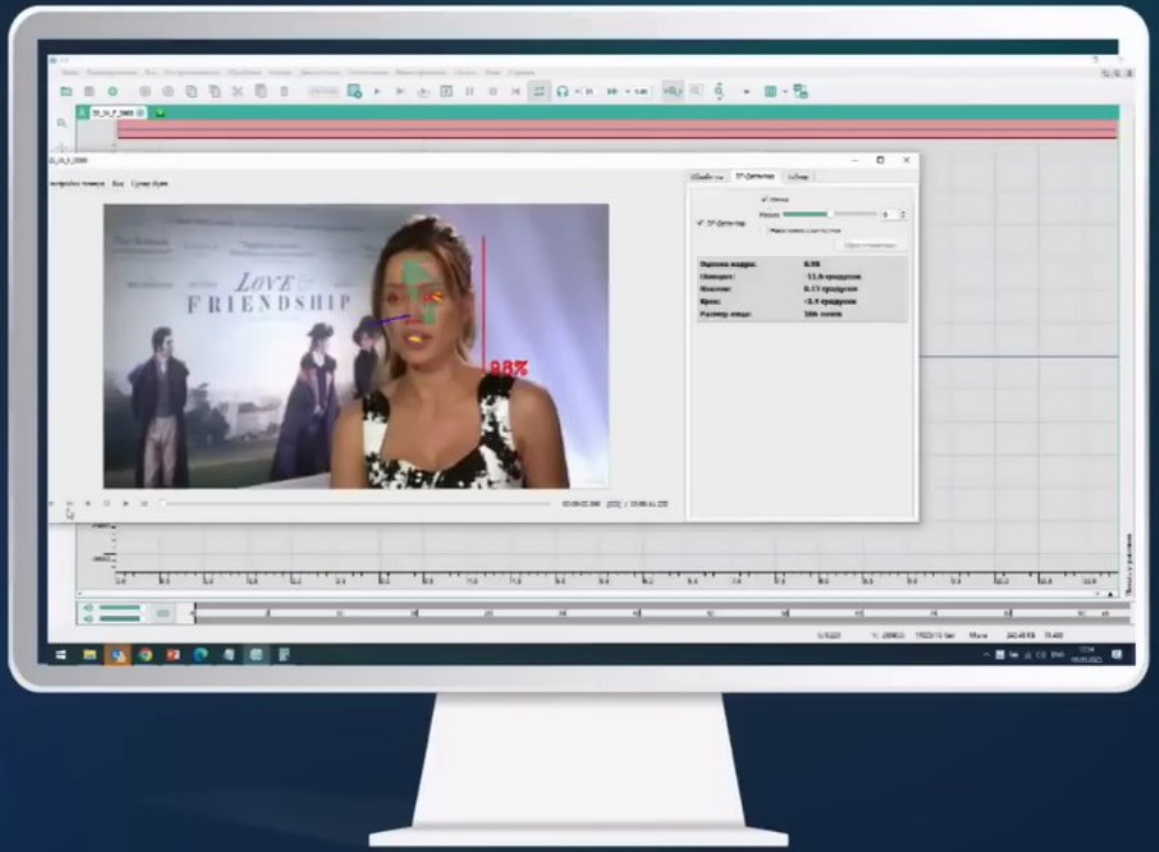
- The video frames of the media file are analysed
- A high probability of a deepfake has been determined for the synthetic video
- Frame-by-frame analysis shows a score for every frame
- After the analysis an averaged high probability that a deepfake is present is displayed



www.speechpro.ru

Deepfake analysis

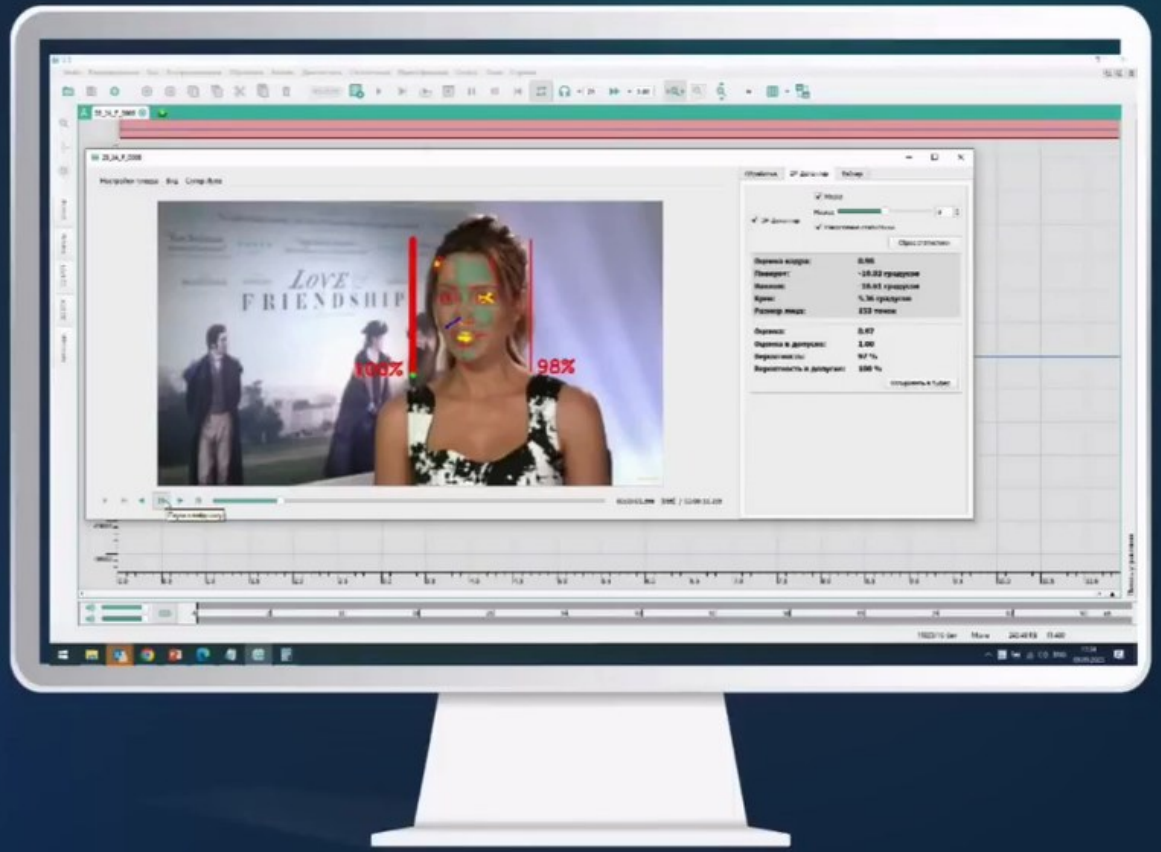
- The video frames of the media file are analysed
- A high probability of a deepfake has been determined for the synthetic video
- Frame-by-frame analysis shows a score for every frame
- After the analysis an averaged high probability that a deepfake is present is displayed



www.speechpro.ru

Deepfake analysis

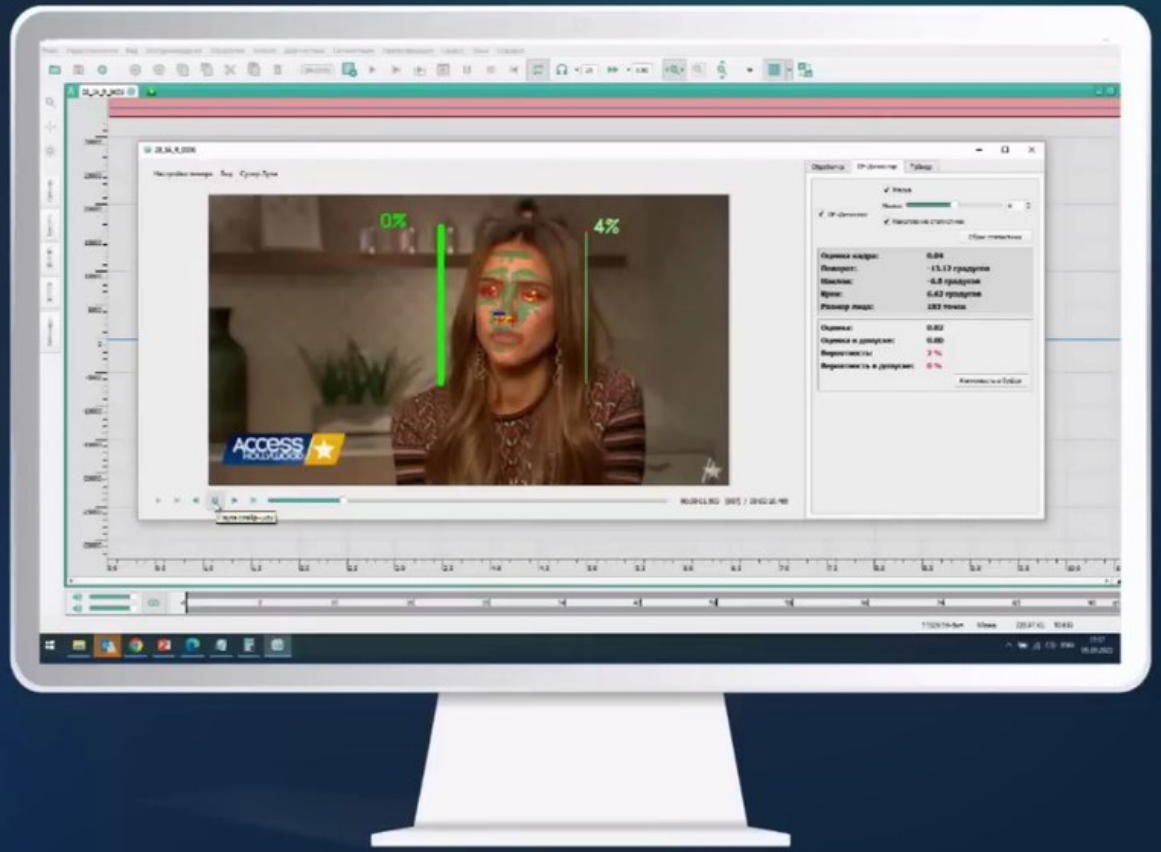
- The video frames of the media file are analysed
- A high probability of a deepfake has been determined for the synthetic video
- Frame-by-frame analysis shows a score for every frame
- After the analysis an averaged high probability that a deepfake is present is displayed



www.speechpro.ru

Deepfake analysis

- The video frames of the media file are analysed
- A low probability of a deepfake has been determined for the original video
- Frame-by-frame analysis shows a score for every frame
- After the analysis an averaged low probability that a deepfake is present is displayed

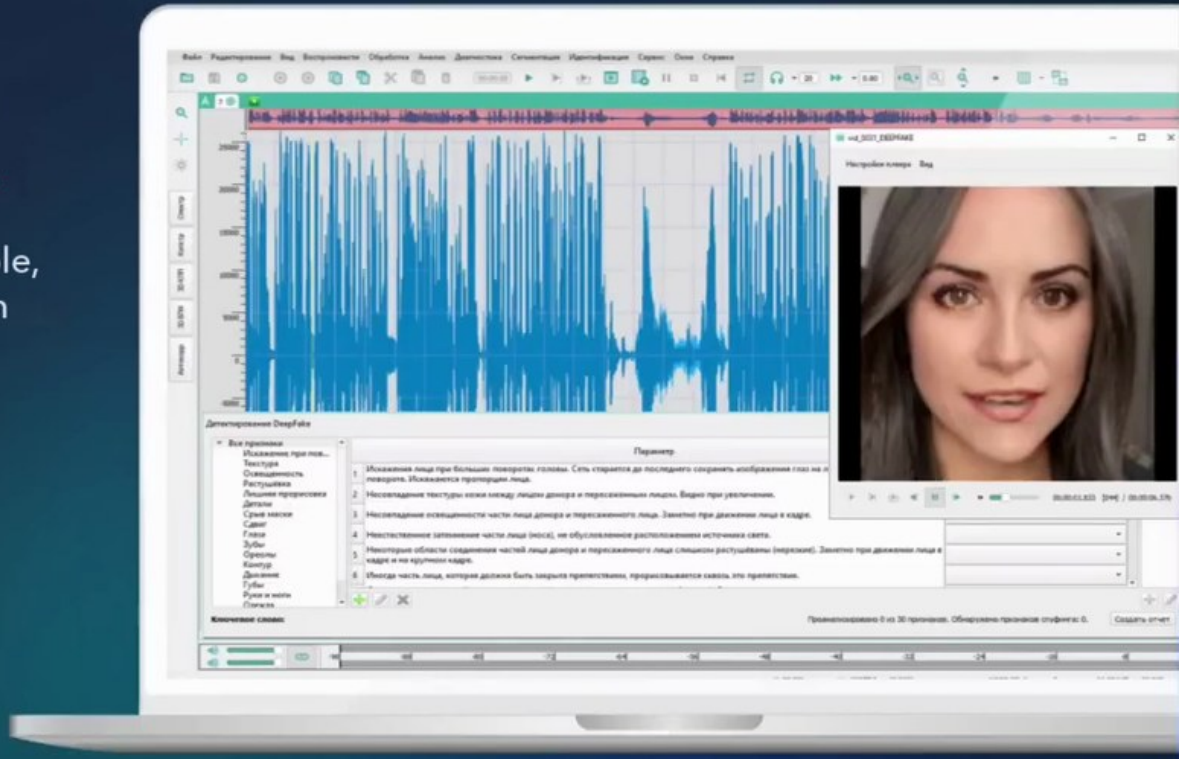


www.speechpro.ru

SIS capabilities

Expert method for detecting deepfakes

- The automatic score can be supplemented by expert analysis of deepfake features* in the video frames
- The expert uses an interactive artefact table, marking the typical distortions not only on the face but across the whole frame
- The result is a report that is saved in the project

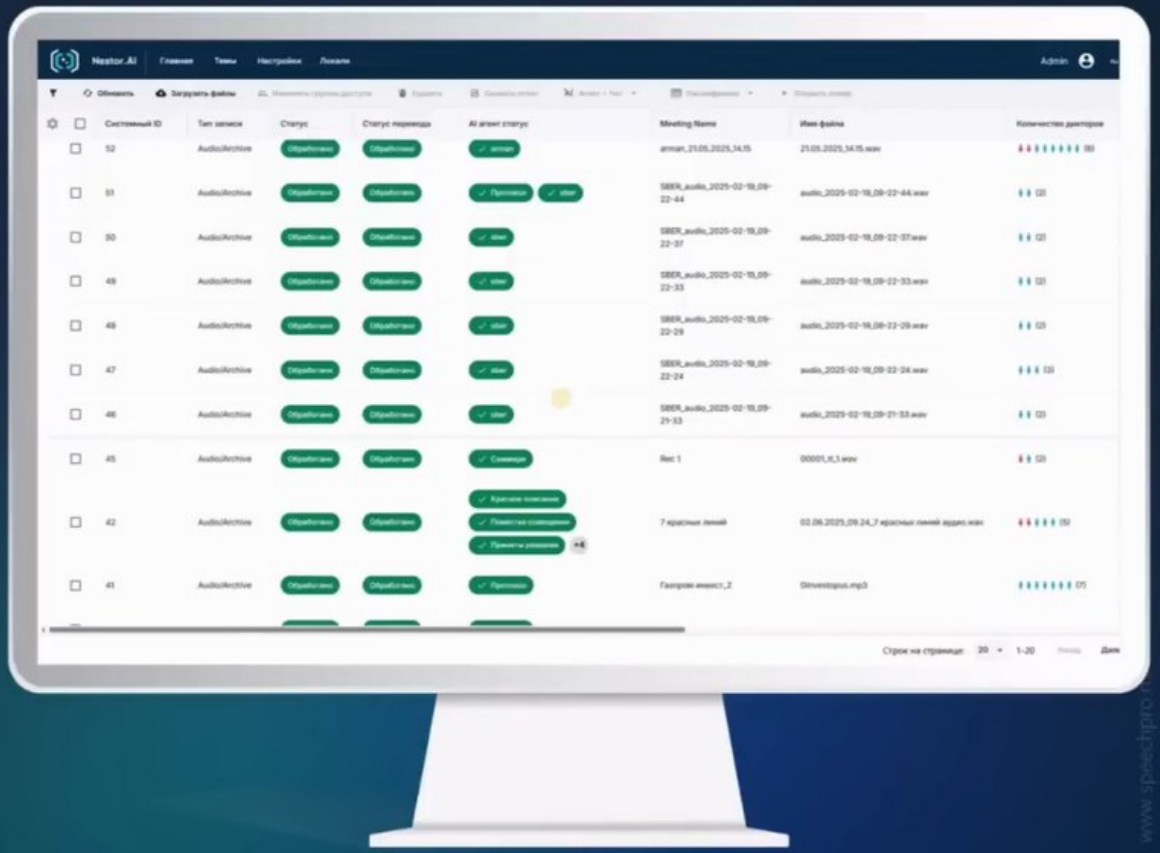


* * Supplied under a separate licence

Nestor.AI

ЦРТ | ГРУППА КОМПАНИЙ

In court proceedings over an audio forgery the Nestor.AI system was used: it automatically turned the speech of the session into text, and the AI assistant prepared a short summary of the content and saved it into a strictly formatted document together with the transcript

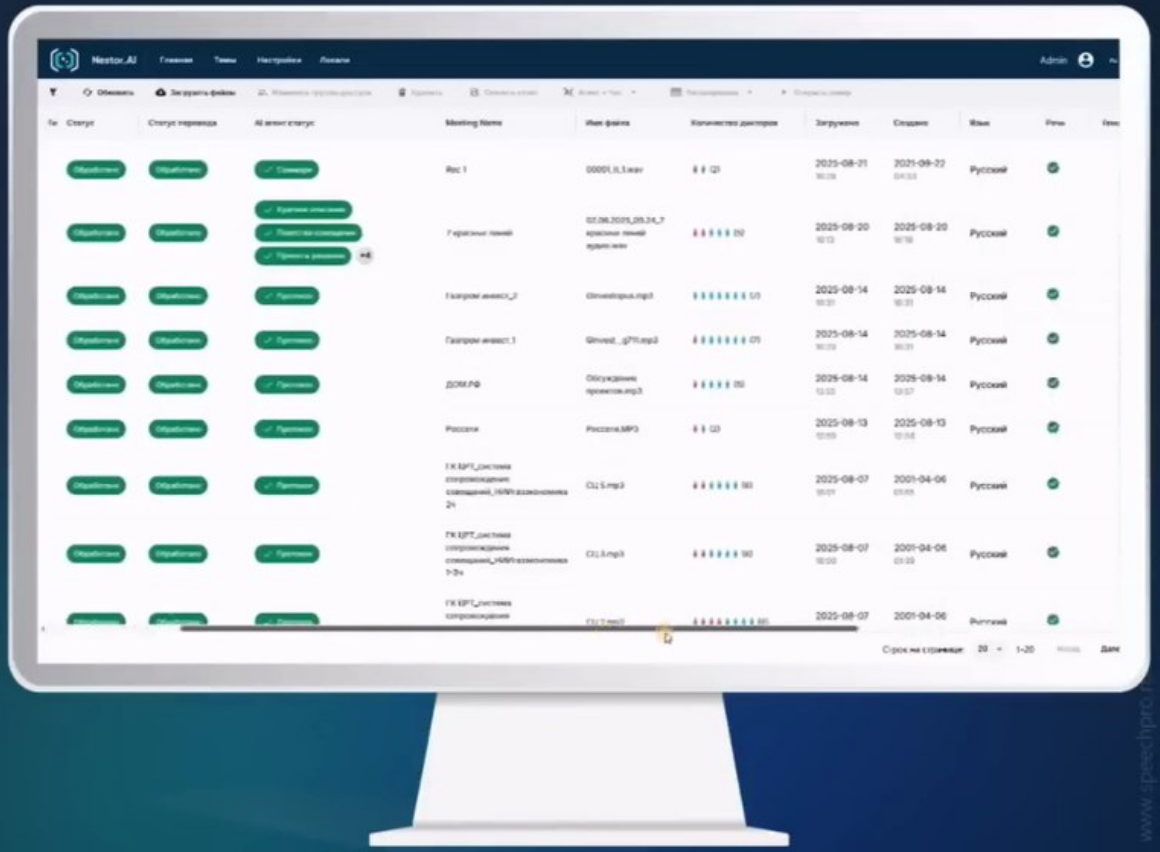


* * Supplied under a separate licence

www.speechpro.com

Nestor.AI

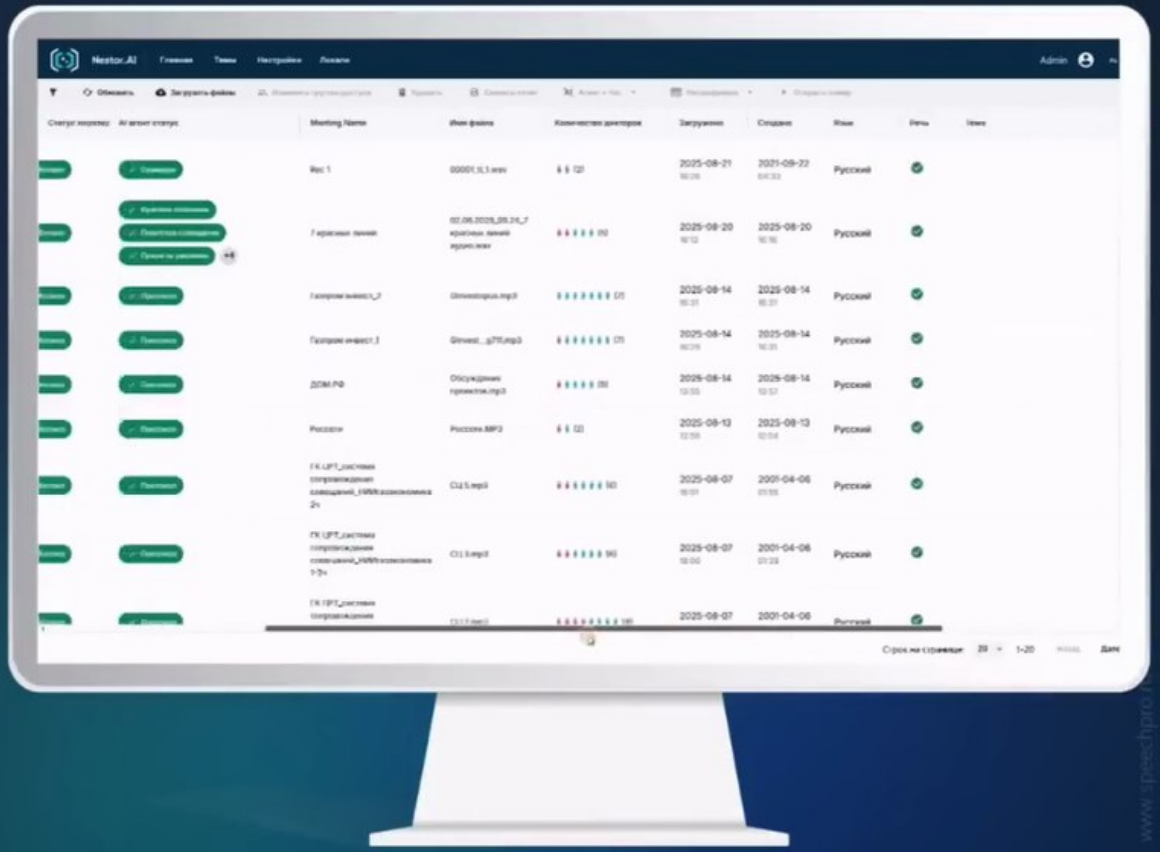
In court proceedings over an audio forgery the Nestor.AI system was used: it automatically turned the speech of the session into text, and the AI assistant prepared a short summary of the content and saved it into a strictly formatted document together with the transcript



* * Supplied under a separate licence

Nestor.AI

In court proceedings over an audio forgery the Nestor.AI system was used: it automatically turned the speech of the session into text, and the AI assistant prepared a short summary of the content and saved it into a strictly formatted document together with the transcript

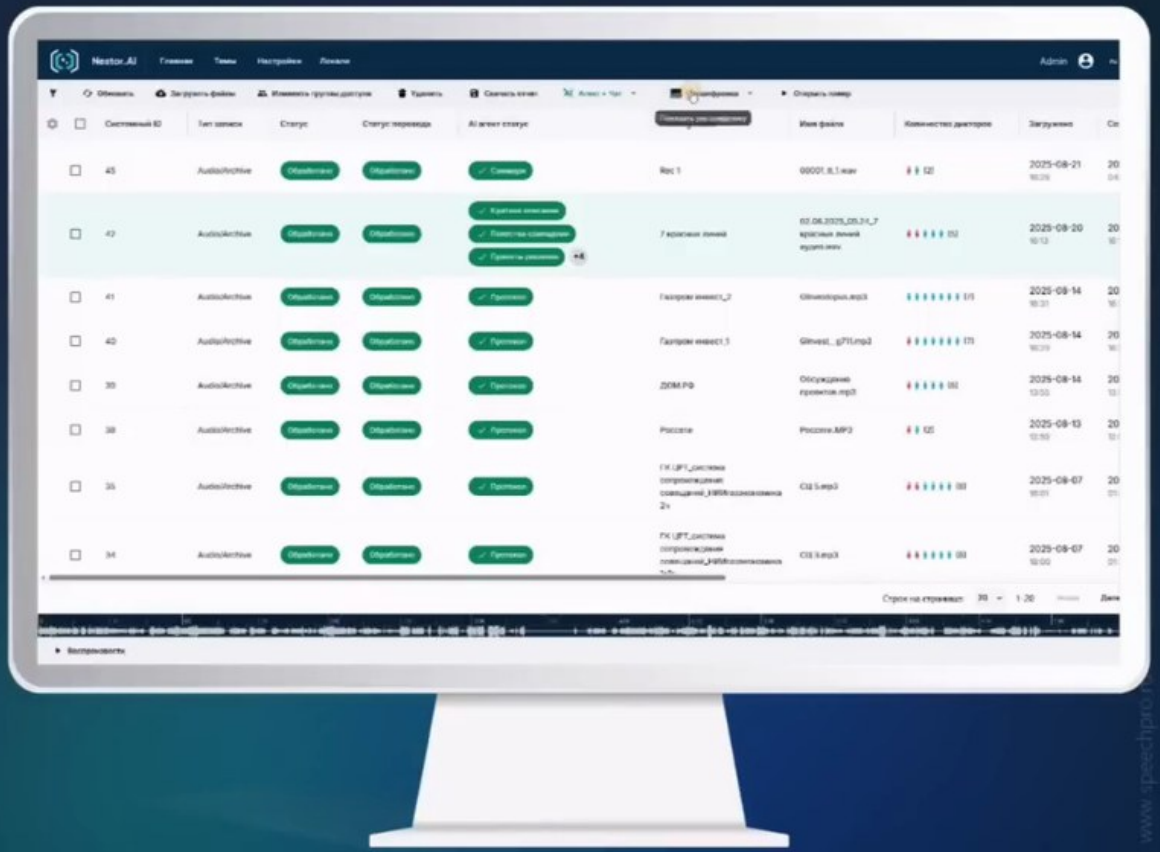


* * Supplied under a separate licence

Nestor.AI

ЦРТ | ГРУППА КОМПАНИЙ

In court proceedings over an audio forgery the Nestor.AI system was used: it automatically turned the speech of the session into text, and the AI assistant prepared a short summary of the content and saved it into a strictly formatted document together with the transcript

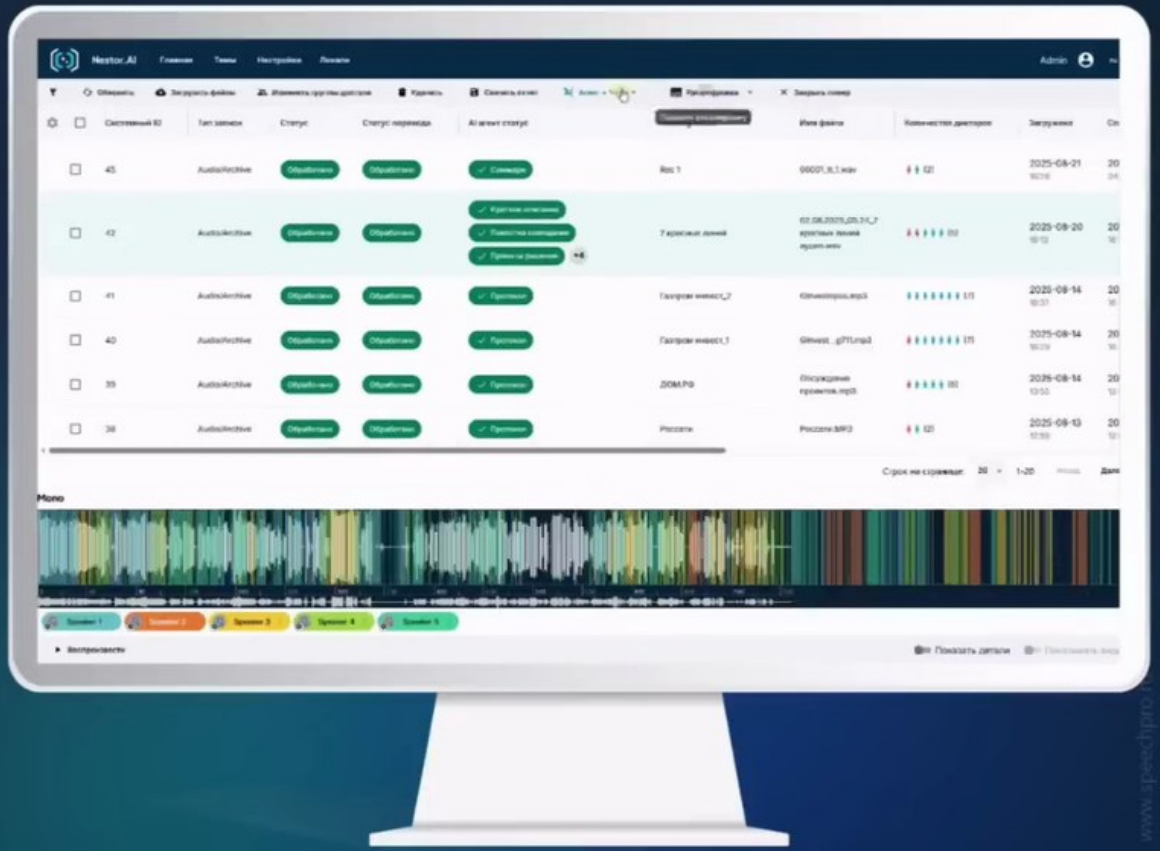


* * Supplied under a separate licence

www.speechpro.ru

Nestor.AI

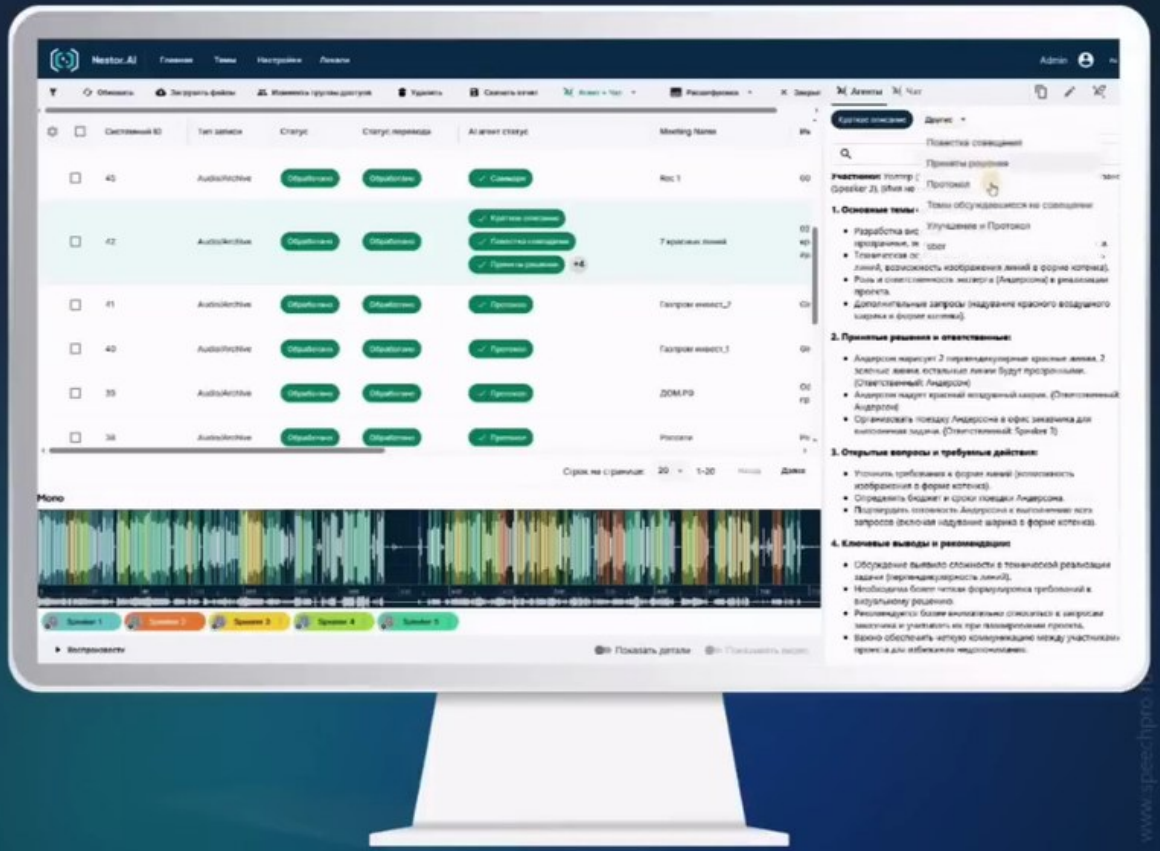
In court proceedings over an audio forgery the Nestor.AI system was used: it automatically turned the speech of the session into text, and the AI assistant prepared a short summary of the content and saved it into a strictly formatted document together with the transcript



* * Supplied under a separate licence

Nestor.AI

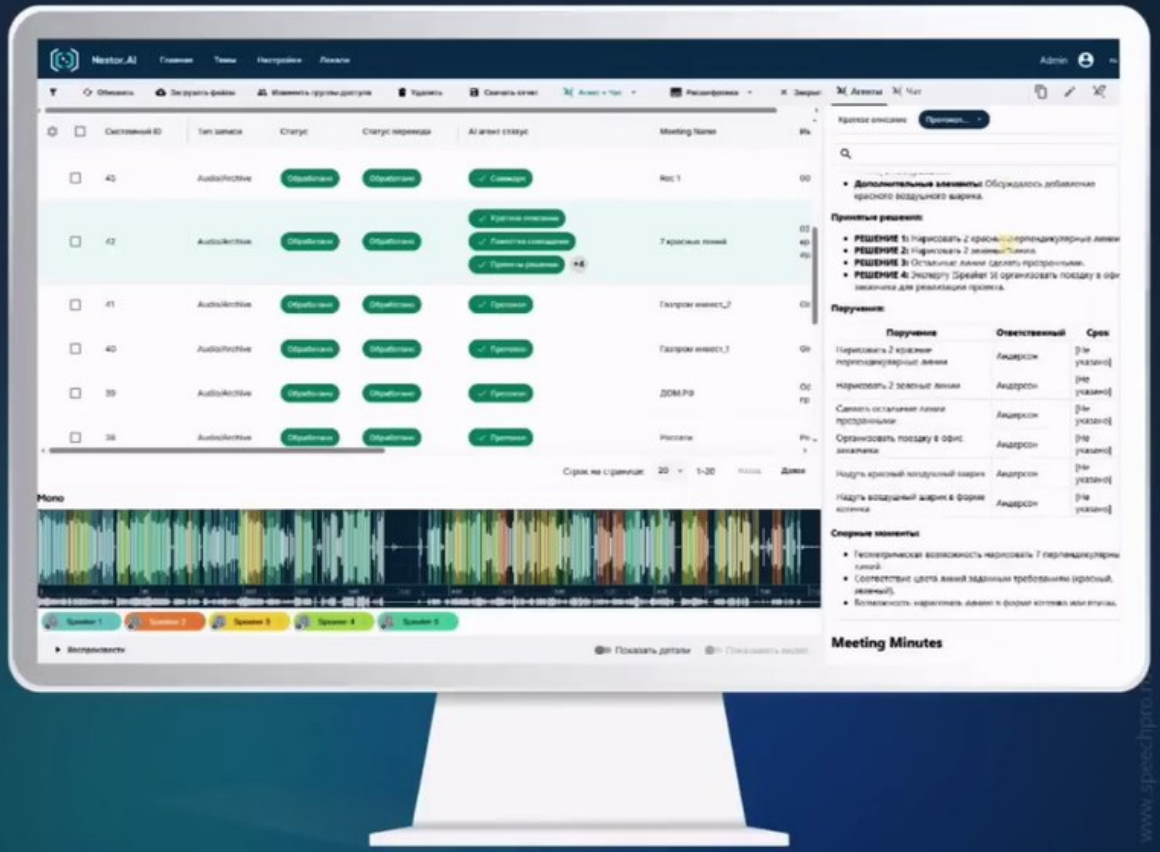
In court proceedings over an audio forgery the Nestor.AI system was used: it automatically turned the speech of the session into text, and the AI assistant prepared a short summary of the content and saved it into a strictly formatted document together with the transcript



* * Supplied under a separate licence

Nestor.AI

In court proceedings over an audio forgery the Nestor.AI system was used: it automatically turned the speech of the session into text, and the AI assistant prepared a short summary of the content and saved it into a strictly formatted document together with the transcript



* * Supplied under a separate licence

Conclusion

The products of the STC group of companies work at different stages of the forensic process and make it possible to use voice biometrics, face biometrics and intelligent speech technologies both to search for the material of interest and to obtain forensic reports

