Description of technology

This technology applies AI algorithms to enable computers to make sense of real-world data, including videos/images (using computer vision), text (using NLP1), or audio (using speech technology).

It also facilitates human–machine interaction and contributes to digitization of natural environments.

By training algorithms with sample data, machines can recognize patterns and interpret and act on them.

What it enables companies to do

1. Seamless human–machine interaction
   Simplifies application usage or interaction with machines by translating human output to machine-readable instructions

2. Extension or creation of new products and services toward end-to-end coverage
   Exploits AI-based data-analysis features (eg, cancer detection as feature of medical-imaging devices)

3. Higher degree of digitization and automation
   Audiovisual comprehension of machines allows companies to fully automate or to aid humans by eliminating multiple tasks

4. More customer interfaces
   Offers different interaction methods to engage with customers either via text, speech, or image (eg, Amazon’s feature to search for products based on photos)

Example use cases

1. Extension or creation of new products and services toward end-to-end coverage
   Gained by exploiting AI-based data-analysis features (eg, cancer detection as feature of medical-imaging devices)

2. Higher productivity of (human) resources
   Gained from efficient human–machine interactions, ie, computer-vision/NLP algorithms facilitate processes and substantiate decision making (eg, by searching contracts for specific information) such that employees spend time on value-adding tasks

3. Full automation of tasks reducing operating expenses
   AI-controlled robots can efficiently process repetitive back- and front-office tasks (eg, autonomous grocery stores using computer vision for checkout)

4. Higher customer satisfaction
   Offers a variety of natural interaction methods at a consistent service level at all times (eg, chat or voice bots)

Technology maturity

By 2024, more than 50% of user touches will be augmented by AI-driven speech, written word, or computer-vision algorithms

By 2021, 1 billion connected surveillance cameras will collect and share visual data

Industry applicability

1. Natural-language processing.

Source: Expert interviews; Forrester predictions on AI (2020); IHS Markit report (2020); McKinsey analysis
Use case deep dive: Computer vision/ NLP/speech tech

Proof points

<table>
<thead>
<tr>
<th>Use case</th>
<th>Situation and approach</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Extension or creation of new products and services toward end-to-end coverage</strong></td>
<td>Crop-protection producer wants users to optimize product usage for plant-disease prevention. As diseases often remain undetected, computer vision is used on live satellite images of fields to trigger alert for risk of infection.</td>
<td>Computer-vision algorithm detects growth problems early before they are visible for human workers. Triggers an alert 14 days in advance of actual infection, allowing farmers to combat viruses with mild dosage of pesticides.</td>
</tr>
<tr>
<td><strong>Higher productivity of resources</strong></td>
<td>UK's Serious Fraud Office has more than 30 million deposited documents. Applied computer vision to sort documents and perform basic data handling to detect corruption cases.</td>
<td>In 2017, it detected Rolls-Royce fraud case, charging the firm a £671 million fine. 600,000 documents processed daily and saving 80% of time for documentation handling, allowing to refocus effort on prosecution.</td>
</tr>
</tbody>
</table>

Source: Expert interviews; McKinsey analysis
Expected technology-development horizons: Computer vision/NLP/speech Tech

Expected technology-development horizons in next 5 years

<table>
<thead>
<tr>
<th>Enablers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Standardization and proliferation of computer vision and NLP models</strong></td>
</tr>
<tr>
<td>Pretraining of language or vision models on a diverse set of unlabeled data allows for context-aware content creation, often indistinguishable from human-created content</td>
</tr>
<tr>
<td>Eliminates need for human supervision or manual labelling of input data</td>
</tr>
<tr>
<td>Follows own logic rather than building on patterns to enable realistic human–machine interactions, allowing for independent “thinking” and creating content with high accuracy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Barriers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Flaws with existing neural networks remain (eg, lack of context) limiting power of interpretation and ability to act or build on insights</strong></td>
</tr>
<tr>
<td><strong>Growing privacy ramifications globally limit data collection and use cases</strong></td>
</tr>
</tbody>
</table>

Source: Expert interviews; McKinsey analysis