



# Data Quality at Toluna



# Market Landscape



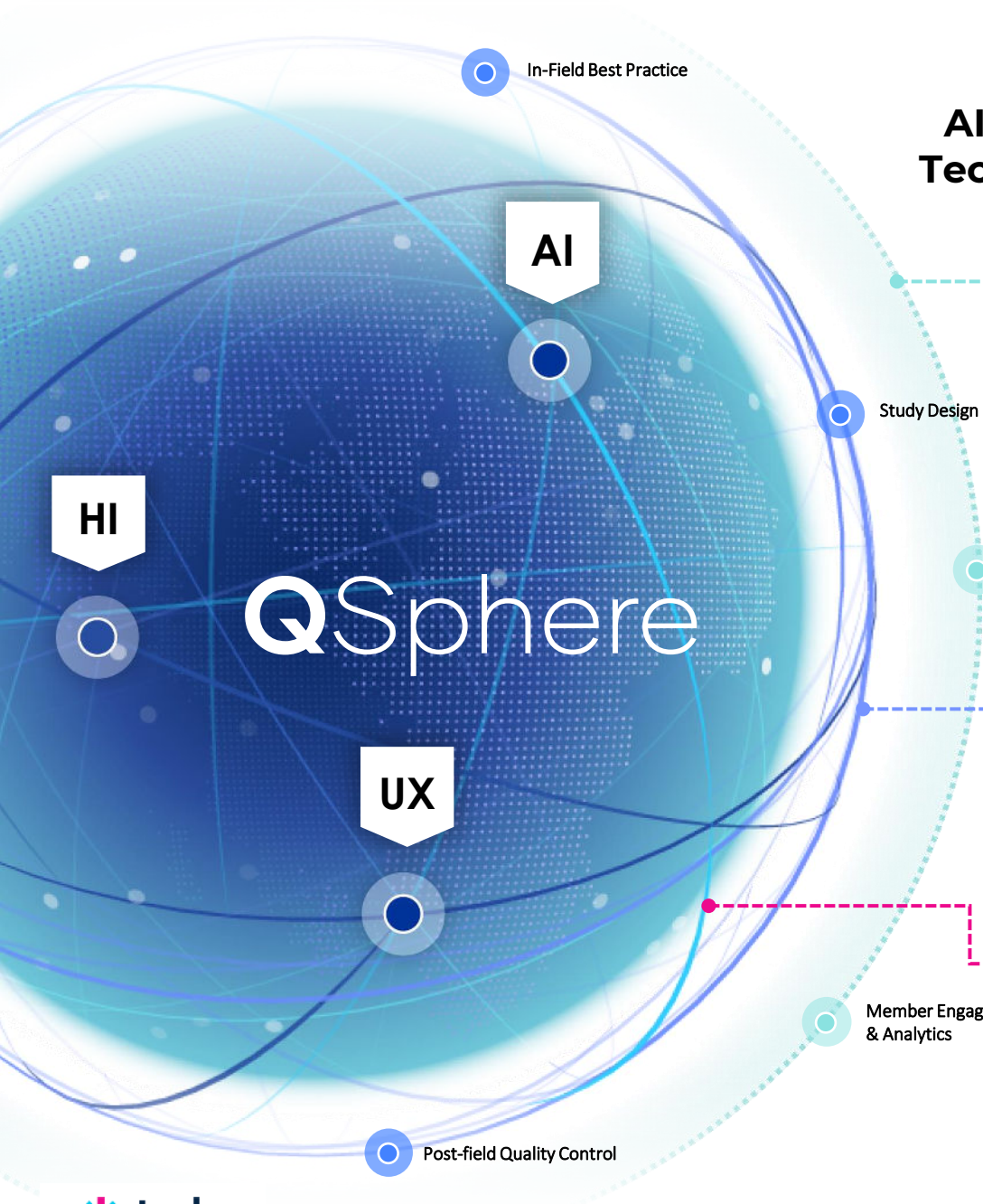
# Quality Challenges in Online Research

## Today's challenges

- ✓ Global, organized fraud
- ✓ Evolution of research best practices
- ✓ Competition for respondents' attention
- ✓ Instability of tracking data
- ✓ Lack of transparency in respondent sourcing and motivation

## Toluna's differentiators

- ✓ **Transparency and longitudinal respondent monitoring** on our proprietary panel, which is the Toluna source with the lowest pre-survey rejection rate.
- ✓ **Combination of research, technology, supply, and analytics expertise.**
- ✓ **Pioneers** in online research with the ability to adapt to rapidly changing industry trends.



**AI-driven  
Technology**



**Human  
Intelligence**



## Continuous Quality Cycle

Integrating human intelligence (HI) and AI across research, respondents, and platform for comprehensive quality measures.

### Respondents

Toluna vets respondents in real-time using AI and machine learning, combining device, behavior, and identity checks. Low-scoring respondents are flagged and restricted from re-entering surveys, ensuring the integrity of our data.



The AI learns from each interaction to improve respondent vetting processes.

### Research

Toluna's research experts and AI-driven tools create engaging, unbiased surveys. Our smart methodologies ensure high-quality data collection. Continuous learning from aggregated data refines our best practice standards.



The AI constantly learns from field data to enhance future survey designs.

### Platform

Our intuitive tools harness best research practices effortlessly. While you may run projects on your platform, our knowledge base is open, championing shared expertise. Automated quality checks and UX tests ensure impeccable data quality.



The AI continuously learns from each project to enhance survey quality and user experience.



# Key initiatives

2023

A focus on **automation**, **collaboration**, and **fraudulent behaviour**.

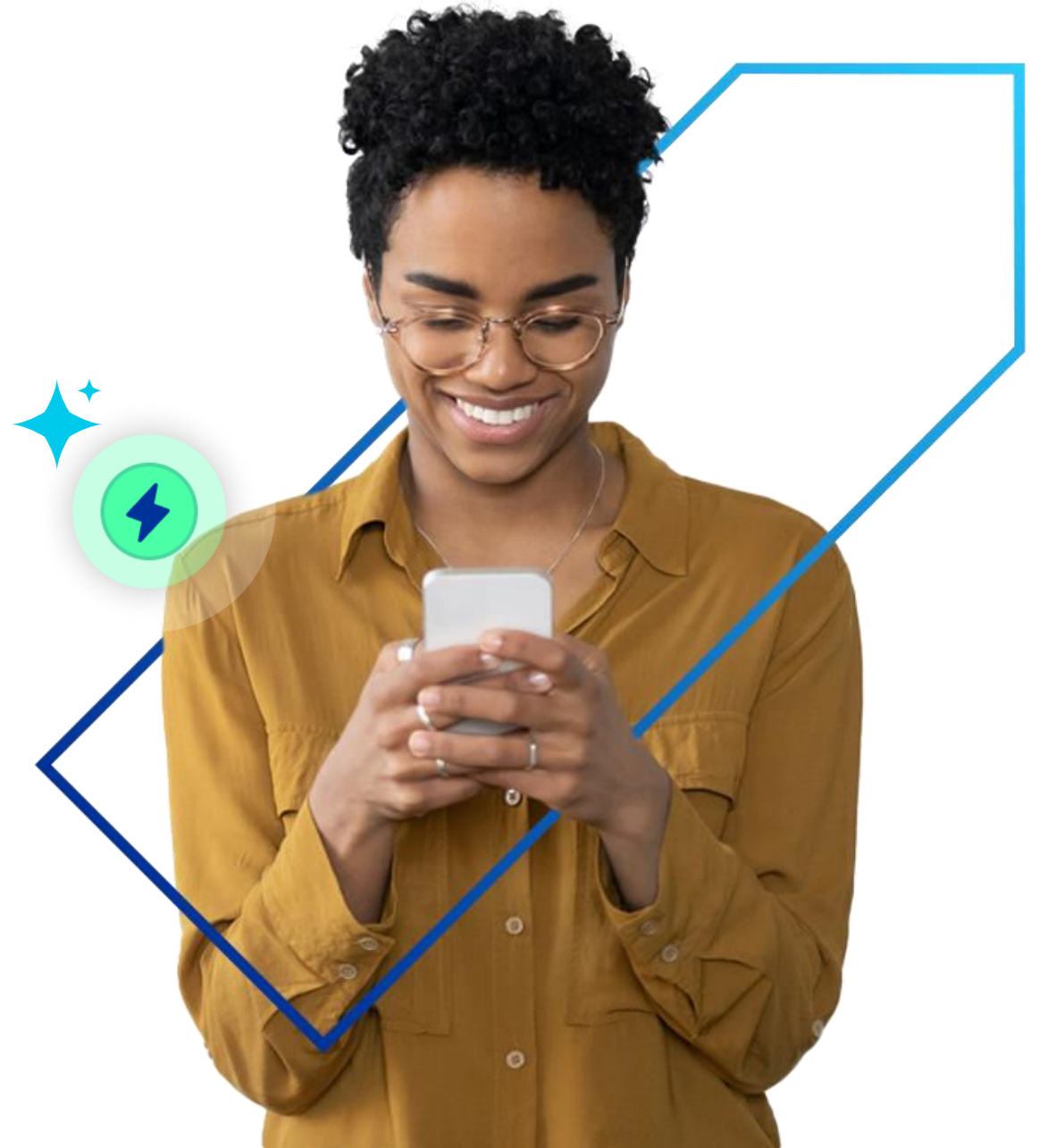
- ✓ **Automation of open-ended checks** through our automatic detection of duplicate, gibberish, or foreign language via our in-house NLP algorithm.
- ✓ Tackling **fraudulent behavior** through copy & paste detection, overstatement, and language comprehension checks
- ✓ **Collaboration** with industry initiatives to develop best practice approaches for methodology and data quality processes.

2024

A focus on **respondent identity**, **proactive automation**, and **future-proofing**.

- ✓ **Exploring biometric verification** as a respondent identity check.
- ✓ **Future-proofing** in- and post-field quality checks against emerging threats like AI and ChatGPT.
- ✓ **Automating our in-field monitoring** to proactively identify projects with potential performance issues.

# Respondent Quality



A decorative graphic on the left side of the slide. It consists of a large, semi-transparent sphere with a blue and green gradient. Inside the sphere, there are several concentric circles and a network of thin, light blue lines connecting various points, resembling a globe or a data visualization. A dashed green line extends from the top of the sphere towards the right, ending near the "Respondents" header.

## Respondents





Our continuous respondent monitoring algorithm collects data on respondents at several stages of the research process: during registration, in pre-survey checks, via automated in-field quality checks, and in post-field cleaning. Leveraging machine-learning and AI, our respondent scoring algorithm combines these data points into a real-time decision on whether a respondent meets our quality expectations or not.

The effectiveness of our quality checks speaks for itself:

**We catch over 30% more survey attempts at the pre-survey stage than only 12 months ago (Q1 '23 vs Q1 '24).**

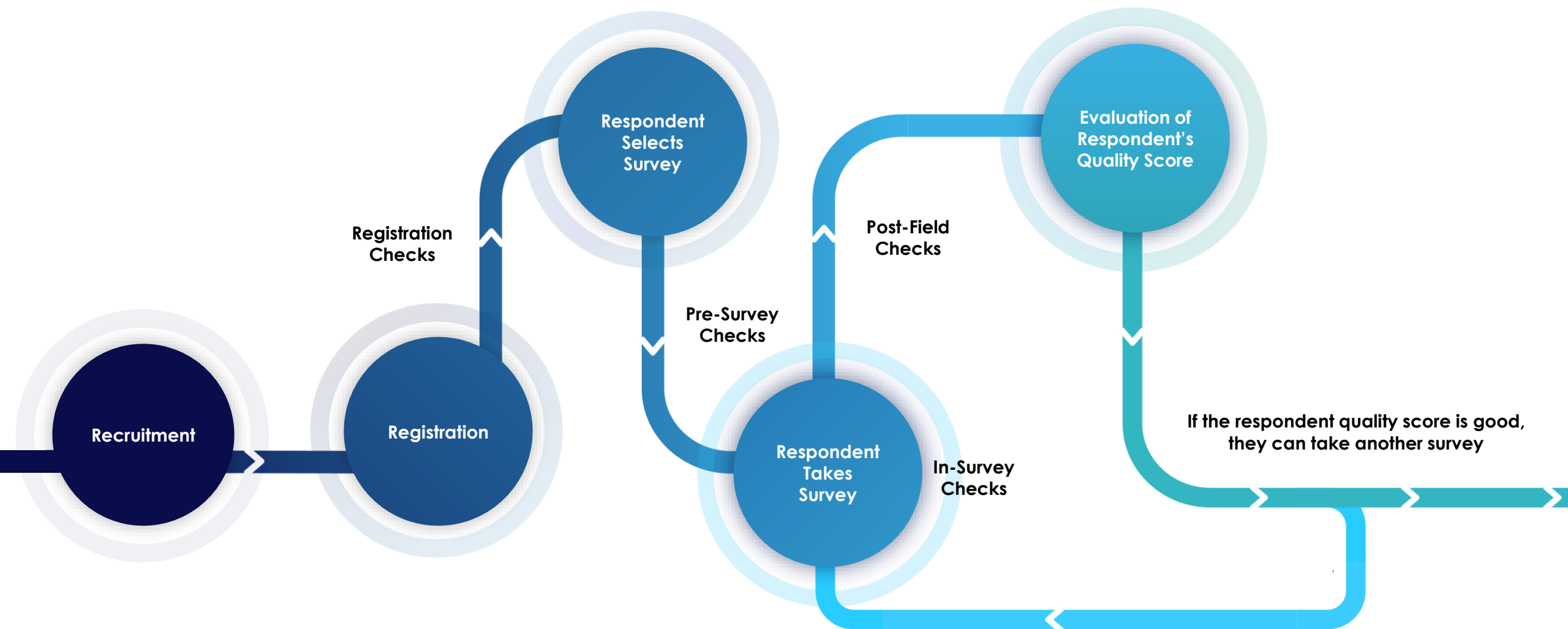
This saves you valuable time at the post-survey data cleaning phase while ensuring the highest quality of data.

Some highlights of our suite of quality checks includes:

-  A multi-faceted approach of combining device, behavior, and identity checks
-  Real-time verification of IP addresses against a cross-industry fraud database
-  AI-focused checks such as copy & paste detection
-  Fraud behavior focus via overstatement detection

# Process

Real-time respondent quality scoring algorithm based on device, behavioral, and identity information





# Objective

Create a respondent base that's real, genuine, engaged, and representative



**We combine three categories of quality checks to achieve optimal data quality.**

## Device

geolocation, language & time zone settings, IP address settings, reCAPTCHA, use of auto translators...

**9.4%**

of survey attempts rejected for quality reasons

## Behavior

overstatement, language comprehension, attention, copy paste, straight-lining, speeding, OE quality...

**1.6%**

of checked open-ends removed by our automated open-end checks

## Identity

digital fingerprinting, respondent PII pattern checks, account grouping, unique and trackable respondent ID...

**5.2%**

of survey attempts blocked for digital fingerprint flags

# Respondent Quality






An abstract graphic on the left side of the slide. It features a large, semi-transparent blue sphere with a grid of white dots and intersecting blue and purple lines. A dashed blue line extends from the top of the sphere towards the right, ending near the "Platform" heading.

## Platform

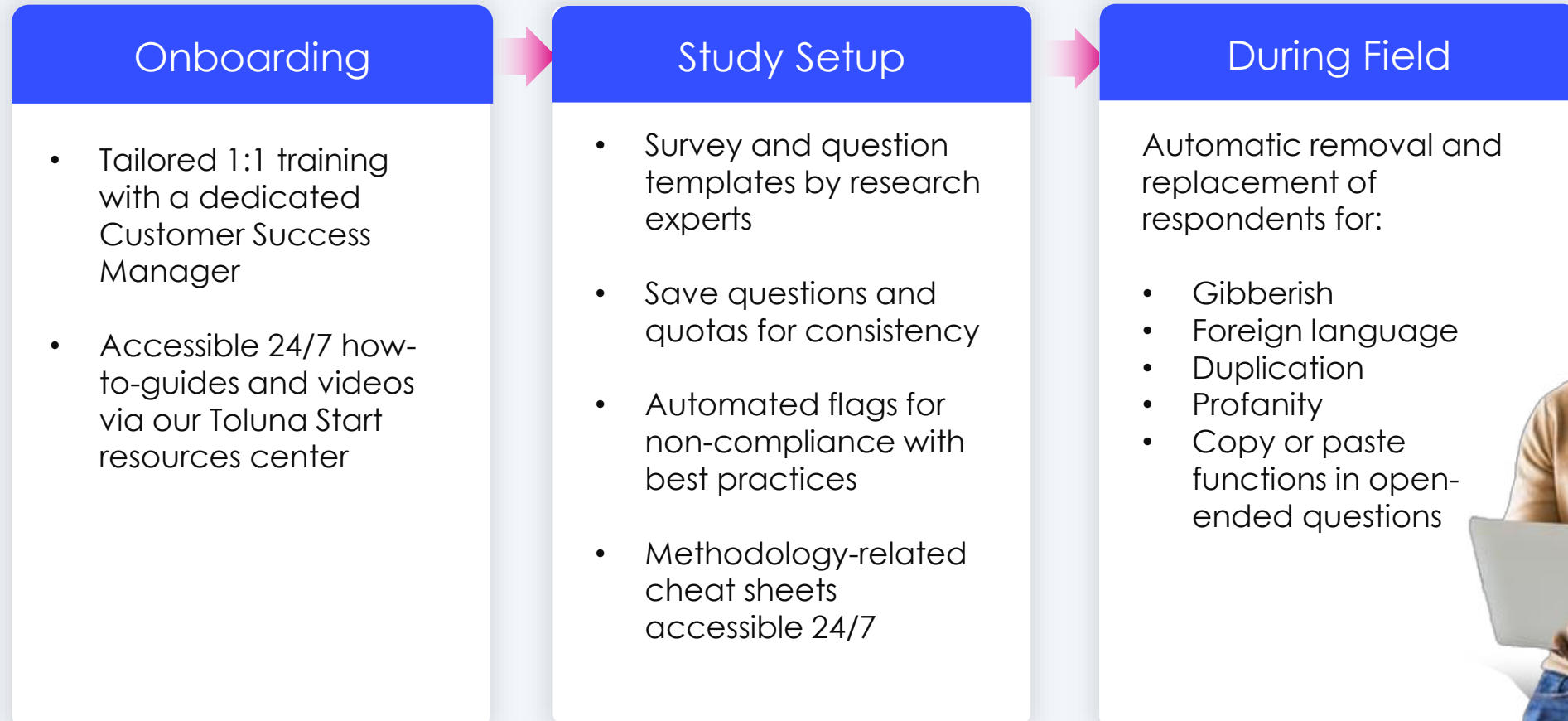
You don't have to run a full-service project with us to feel like you have an expert thinking along on your projects. Our platforms are built by researchers for researchers and thanks to our suite of support options and built-in nudges, our platform will guide you in your research design.

Our smart platform also allows you to dive deeper into your insights by leveraging AI for insights, not just data. And you can rest assured that your survey's quality is assured thanks to our automated algorithms seamlessly facilitating your cleaning process in the background.

### **Some highlights of our research approach include:**

-  Get more insights from open-ends thanks to our AI-powered SmartCloud tool
-  How-to guides, cheat sheets, and videos accessible 24/7
-  Automated data cleaning thanks to our automated open-end cleaning algorithm & copy paste detection

# Personalized support throughout every stage of the research process





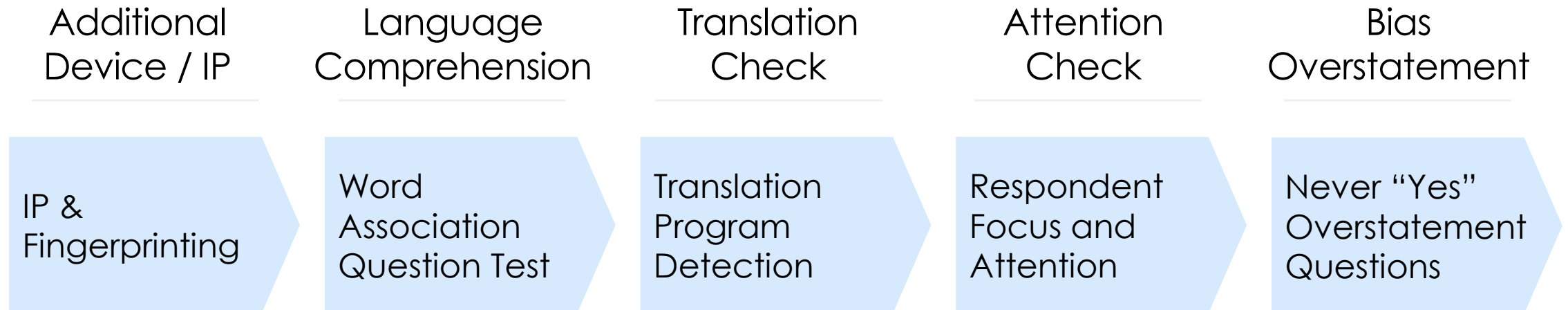


# Sentry Pre-Survey Behavioral Screening

---



## Pre-Survey Behavioral Screening



## Pre-Survey Behavioral Screening

### Language Comprehension

Word  
Association  
Question Test

**Moment is most  
associated with?**

Teaspoon  
**Time**  
Nonprofit

**Statue is most  
associated with?**

Pal  
Testing  
**Monument**

### Bias Overstatement

Never “Yes”  
Overstatement  
Questions

From memory, can you  
recall the name of  
every person who ever  
lived in the country you  
grew up in?

Do you know what the  
word “hufre” means?



# Thank you. ♦

**Copyright © Toluna**

All rights reserved. Nothing from this report is allowed to be multiplied, to be stored in an automated file, or to be made public electronically, mechanical, by photocopies, recording or any other manner, without written consent of Toluna.

