choreograph
A WPP Media Brand

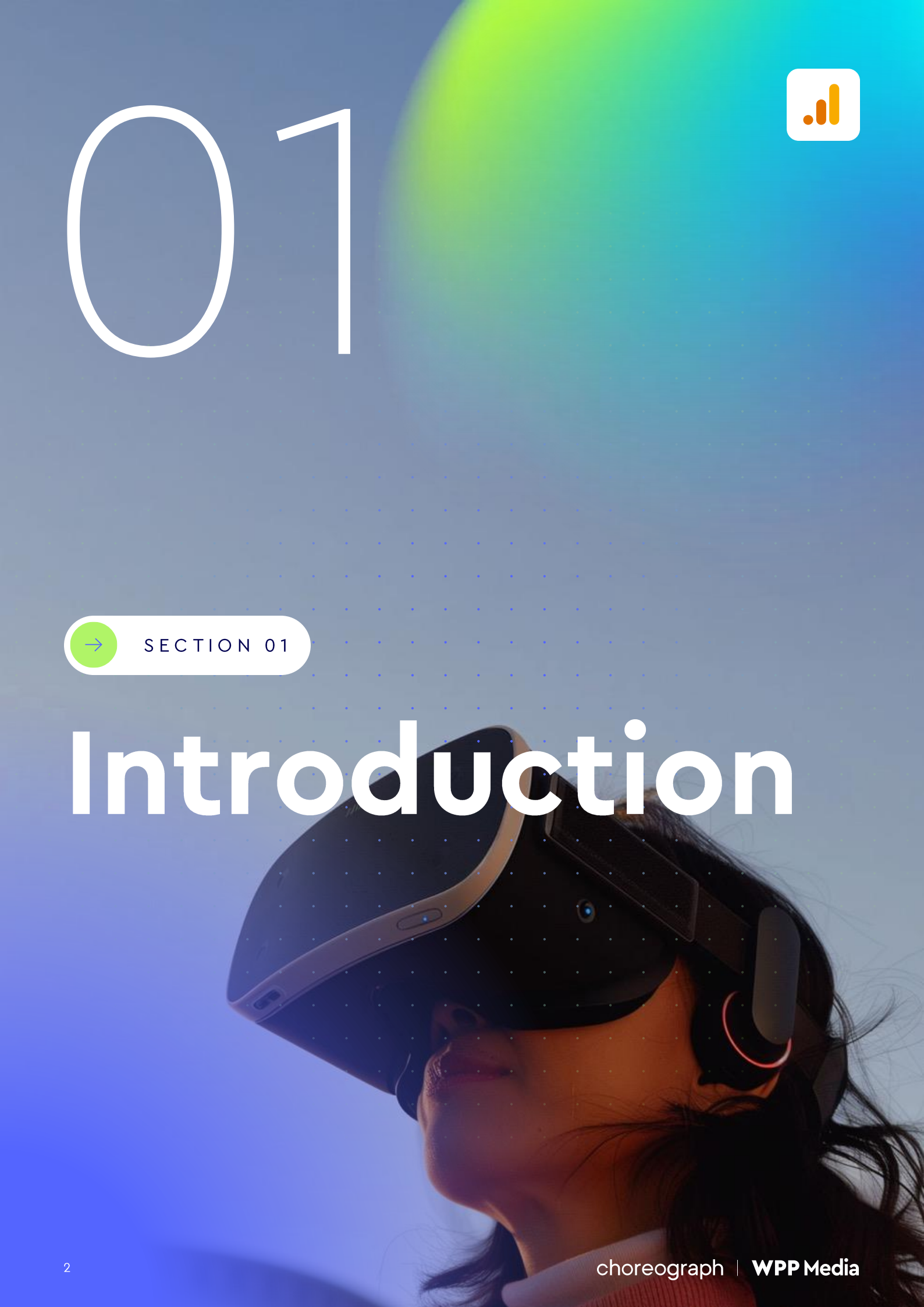
How to Maximize High-Quality, AI-Ready Data in Google Analytics 360

Hollis Ross
Digital Lead | North America

01

SECTION 01

Introduction



Machine learning (ML) and artificial intelligence (AI) are revolutionizing marketers' ability to unlock deeper insights, make strategic decisions, and execute more impactful media and marketing campaigns at scale.

But AI, and ML more generally, are only as good as the data that is fed to them, highlighting the foundational importance of data capture, design, and governance for organizations using AI/ML.

QUALITY AND WELL-DESIGNED DATA CAN:

Build a foundation for accuracy & trust

Capturing clean, consistent, and accurate data ensures reliable results that boost trust and adoption.

Reduce downstream costs

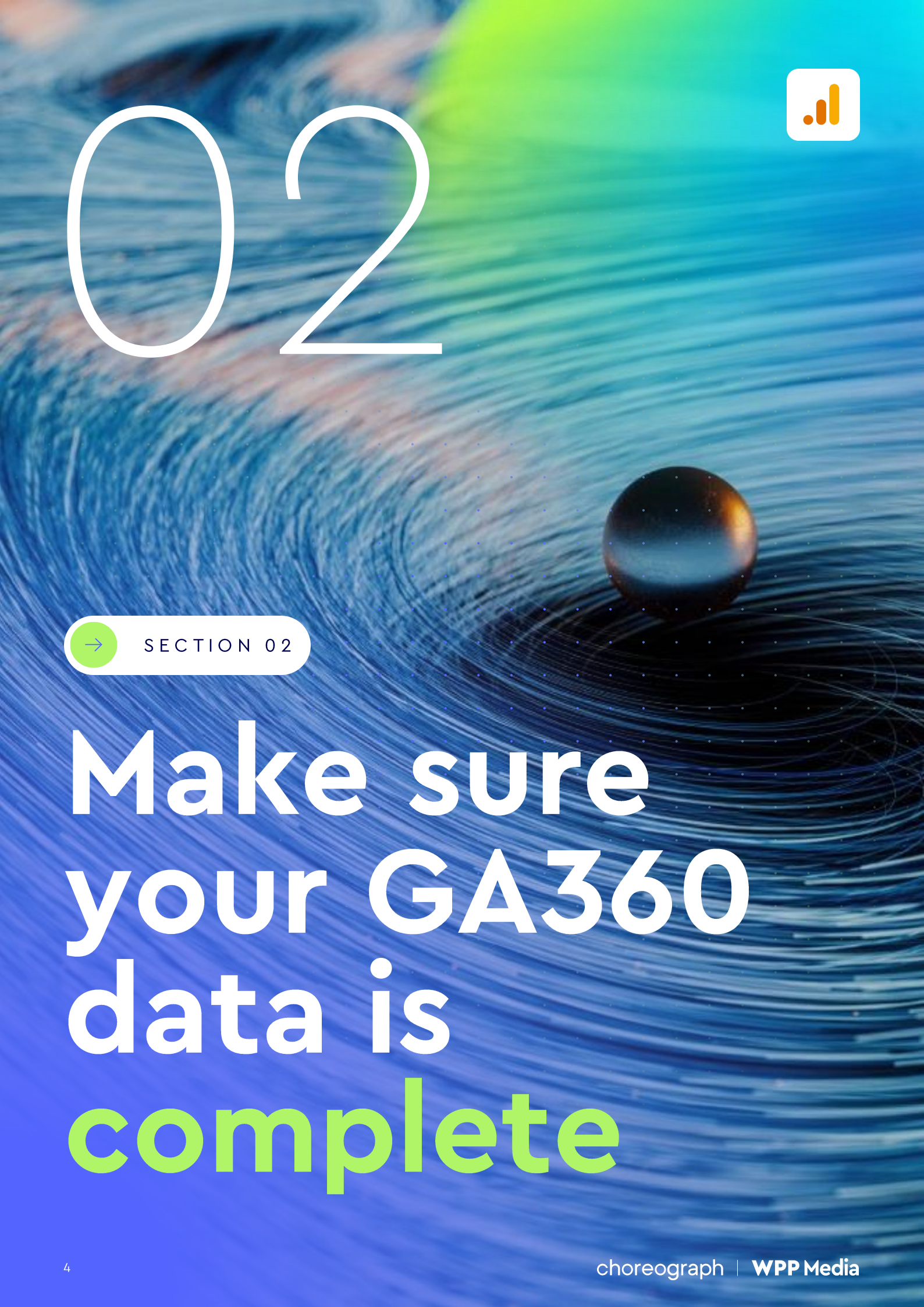
Poor data capture leads to costly rework, including cleaning and potential retraining of models.

Increase speed to value

By increasing efficiency and minimizing rework, high-quality data enables more rapid development and deployment of AI solutions.

For many businesses, one of the richest touchpoints for consented and high-quality data capture at scale is their website. **That's where this guide comes in.**

With a focus on Google Analytics 360 (GA360), we'll share how marketers can configure GA360 to capture AI-ready 1PD from their websites. You'll learn how to make sure your GA360 deployment is tuned to collect 1PD that is complete, accurate, representative, and connected, empowering you to fully capitalize on the power of AI.



02

SECTION 02

Make sure your GA360 data is complete

Data completeness in ML refers to the **degree to which all relevant data points and features are present in a dataset.** A complete dataset has minimal or no missing values for the variables being considered.

Incomplete data can lead to inaccurate predictions, as the model doesn't have the full context or enough examples to learn from. Missing data can skew results and introduce bias, especially if certain groups or behaviors are underrepresented. Complete datasets help models learn patterns that generalize better to new, unseen data, reducing overfitting to partial or skewed inputs.

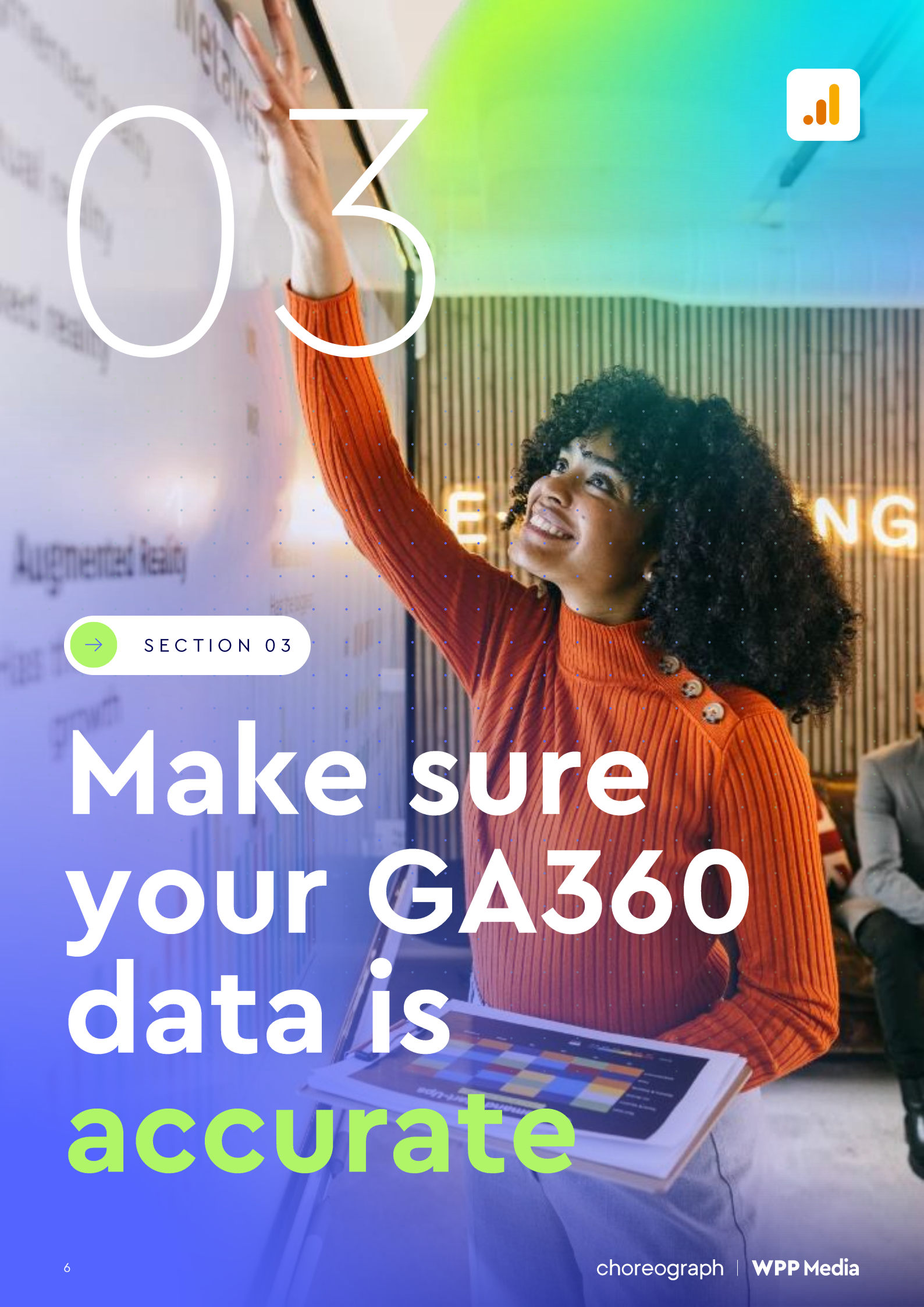
HOW TO ENSURE DATA COMPLETENESS IN GA360:

1. **Use a data layer and tag manager, backed by a comprehensive tagging strategy**
Structured, comprehensive data collection starts with a strong data strategy. Defining what to track and how each element connects to lay the groundwork for an effective data layer and tagging implementation. Leveraging these tools will also minimize site load times, maximizing the volume of data collected and ensuring consistent tagging across all areas of the site. Ultimately, this guarantees that all meaningful data is collected and available for use with AI/ML.

2. **Implement Enhanced Conversions**
Sending hashed customer information helps connect conversions to media and online activity, enabling the development of more accurate, comprehensive models and user journeys for your AI applications. Learn more about Enhanced Conversions here.

3. **Collect as many opted-in identifiers as possible**
Maximizing the number of stable, consented identifiers such as user IDs and user-provided data in GA enables more accurate cross-device and multi-touch attribution. It also improves your ability to align as much data as possible with enriched data from models and 3P sources. The result: more relevant, user-scoped signals that enhance the performance and precision of your AI applications.

4. **Use BigQuery Exports and upgrade to GA360 to avoid volume caps and sampling**
For AI readiness, access to granular data is key. One effective approach is connecting GA360 with BigQuery (BQ) to push your analytics data into an owned cloud data warehouse. From there, you can easily integrate with ML tools such as iBQML, BQML, Vertex AI, and more. Using BQ with GA360 eliminates limitations such as BigQuery download caps or in-UI sampling, ensuring significant training and validation datasets for modeling.



03

SECTION 03

Make sure your GA360 data is accurate

In AI/ML, data accuracy refers to the **degree to which the data is correct, truthful, and free from errors, inconsistencies, and biases.** Accurate data reflects the true state of what it represents.

Inaccurate data leads to poor predictions and unreliable outputs because the model learns from incorrect information. AI systems trained on inaccurate data can amplify mistakes, especially in real-time or automated decisions. Inaccuracies can distort real-world patterns and lead to biased or unfair outcomes.

HOW TO ENSURE DATA ACCURACY IN GA360:

1

Perform comprehensive QA
Data accuracy is built on great QA. As part of the deployment cycle for GA360 events, check data thoroughly in lower environments and in production once it's live. Check for the expected and anything unexpected. Use GA360 reports and debugging tools to verify implementation. After major site changes, check measurement stability. Audit triggers and ensure best practices are applied when creating triggers to prevent fragility (such as jQuery-based triggers breaking when a form ID is updated).

2

Label your n/a data clearly
When values are blank because no data exists, treat these with a 'null' label to ensure it's obvious that the lack of data is not an error. True blanks in data should be an indication of incomplete tracking and should be addressed.

3

Ensure consistent labels for data quality and AI readiness
To feed AI models with the best quality data, it's vital for the machine to interpret data correctly. Variations like "Facebook," "fb," and "FB" can confuse models, so standardize labels across your dataset. To save yourself time and effort with data cleaning, collect data in GA360 using consistent capitalization, spelling, and, for global sites, a single centralized language for labels. As part of your broader data capture strategy, establish a process to ensure your data labelling rules are respected and followed long-term.

4

Normalize with other sources
To ensure your implementation is working as expected, periodically compare with other trusted sources. You'll never get an exact 1:1 match in data between different source collection mechanisms (e.g. GA vs. ad platforms like social media, GMP, and Google Ads; or your order management system), but data should be consistent over time and hopefully within +/-10%. Take a directional approach and if something is significantly off, it could indicate a GA issue or an issue with another system that GA has simply uncovered.

04

SECTION 04

Make sure your GA360 data is representative

In AI/ML, data being **representative** means that the dataset accurately reflects the characteristics and distribution of the broader population or the real-world scenario that the model is intended to learn from and make predictions about. A representative dataset captures the diversity and proportions of different subgroups and features present in the overall population and is important for a model to be able to generalize on unseen data.

Models trained on representative data perform better on real-world, diverse scenarios — not just on the subset they were trained on.

HOW TO ENSURE YOUR GA360 DATA IS REPRESENTATIVE:

1

Enable Advanced Consent Mode
User consent options are essential for privacy compliance regulations, but Google's Advanced Consent Mode also generates modeled traffic and conversions in UI reporting to address gaps. Similarly, synthetic traffic and conversion data can help fill in for lost signals, providing additional data for AI applications where signal lost has been significant. This can help with data being representative by providing more signals for AI models to learn from, ultimately enhancing performance.

2

Ensure all media and unpaid traffic (e.g., social, affiliates) are tracked with UTMs
It's vital that all traffic that you send to your site is tagged with information that aligns with your overall data capture strategy. This is often overlooked for organic traffic through sources like social media posts and QR code landing pages. Ensuring the maximum coverage of UTM tracking allows you to get the most comprehensive possible view of your data for AI models, avoiding bias on paid traffic simply because it has more information associated to it.

3

Use GA360's flexible data model to track apps, POS, IoT, and more
One of GA360's latest innovations allows you to collect data not only from websites, but from connected devices, in-store data with measurement protocol, and native apps with GA's mobile SDK for both Android and iOS. This enables you to provide AI models with a more complete view of user interactions across your products and services.

05

SECTION 05

Make sure your GA360 data is connected

Connected data refers to datasets that are **integrated and linked across different systems, sources, or domains**, allowing AI and ML models to access a more complete, contextual, and unified view of information. If data aren't interconnected and interoperable, you end up with data silos that may be useful in principle, but difficult to scale. For AI/ML, more connected data means you can leverage the full context for better feature engineering, cross-domain insight, and scalability across the enterprise. Connectivity in your data is also powerful reinforcement for the values of being complete, accurate, and representative.

HOW TO ENSURE YOUR GA360 DATA IS CONNECTED AND INTEROPERABLE:

1. **Work together**
   Build a cross-functional team, including IT, Marketing, and Executive stakeholders, to align on strategy, avoid blind spots, and ensure all relevant insights and needs are represented in your data approach. This collaboration also supports stronger enterprise buy-in and adoption.

2. **Map user identifiers to GA data**
   Locate user identifiers, such as emails or user IDs, in other systems and explore how they can be connected or mapped to GA data to enrich tracking and attribution. See page 5 for ideas.

3. **Integrate GA with data resources**
   Connect GA with platforms such as your CRM, CDP, or audience platform to gain deeper, unified insights into user behavior, lifecycle stages, and conversion paths. This integration provides larger, richer datasets for your ML models to train on, helping uncover new insights.

4. **Connect with GMP**
   Link GA with GMP tools to enhance audience targeting, campaign performance tracking, and media efficiency across your marketing stack. This integration helps capture the most comprehensive data in GA and unlocks well-organized data for your ML models to utilize effectively.

5. **Use consistent labeling across systems**
   Standardize labels for campaigns, audiences, and other data segments across GA and internal systems to ensure clarity, alignment, and smoother data integration. There's a subtle but important distinction between consistency and accuracy - consistency ensures that the same terms and language are used across systems, not just within one. When systems share related data, using unified labels helps ML models recognize relationships more effectively, reduces computational complexity, and lowers the risk of errors or hallucinations.

06 In conclusion

Successfully harnessing the power of AI/ML in marketing begins with your data — and that journey starts at the source. By configuring GA360 to capture 1PD that is **complete, accurate, representative,** and **connected**; marketers can create a strong foundation for scalable, high-impact AI/ML initiatives.

This guide has outlined not just the "what" but the "how" of configuring your GA360 implementation to ensure the data you collect is truly AI-ready. From leveraging Enhanced Conversions and cross-device identifiers to standardizing labels and integrating across systems, each action helps build a smarter, more future-proof data ecosystem.

With thoughtful implementation and ongoing governance, your GA360 data won't just inform your next campaign — it will drive a more intelligent and efficient marketing operation for the long-haul.

MORE RESOURCES

Choreograph:

- Building a First-Party Data Foundation: Highlights from Rethink ROI
- GMP's Superpower: Seamless Integrations for Scalable Impact
- Orchestrating your 1PD for Long-Term Growth

Google:

- Best Practices: Implementing ML in GCP
- GA4: Setting Up Predictive Audiences
- GA4: Behavioral Modeling for Consent Mode

Get in touch

Want to learn more about how to maximize GA360 data quality?

Email us at **GMP@Choreograph.com** to learn how we can help you unlock more value from your data and Google technology.

choreograph.com