

Instalily GTM Engineer Case Study: DuPont Tedlar AI Lead Generation Prototype

*This document describes a **toy prototype** AI-powered sales lead generation system developed for DuPont Tedlar's Graphics & Signage team. The prototype was built under tight time constraints to demonstrate concept and data flow, so many features are simplified or use synthetic data placeholders, with significant room for improvement in a full-scale implementation.*

Github repo: <https://github.com/Kataraduo/instalily-gtm-case-study>

AI Agent Workflow (Event to Outreach)

The system operates as an **AI agent** that automates the pipeline from gathering industry event data all the way to generating a personalized outreach message for prospective leads. At a high level, the workflow proceeds through the following stages:

1. **Event Data Scraping:** Collect **industry events** relevant to graphics and signage (e.g. [ISA Sign Expo 2025](#)) and retrieve **exhibitor companies** from those events. These events serve as a primary source of potential customers (exhibiting companies that operate in target markets).
2. **Association Data Scraping:** Gather information on key **industry associations** (e.g. International Sign Association, SGIA) which might list member companies or resources. Associations indicate active players in the industry and can supplement the event leads.
3. **Company Lead Compilation:** Compile a combined list of **target companies** from event exhibitor lists and any association sources. This yields a raw lead list of companies potentially interested in DuPont Tedlar's signage materials.
4. **Data Enrichment & Lead Scoring:** Enrich each company's profile with additional context – such as industry classification, products/services, materials used, company size, etc. – and compute a **relevance score** or tier based on how well the company fits DuPont Tedlar's ideal customer profile. This step effectively **qualifies and ranks** the leads.
5. **Stakeholder Identification:** For each high-priority company, identify key **stakeholders** (e.g. executives or directors in relevant roles). For future work, external APIs (like LinkedIn or Clay) can be used to find real contacts. Now the prototype simulates stakeholders with representative job titles. This yields specific person-level contacts at each target company.
6. **Personalized Outreach Generation (TBD):** Finally, for each stakeholder, generate a **tailored outreach message**. The message is personalized with the stakeholder's name and references the company's specific context (from the enriched data) to highlight how DuPont Tedlar's solutions can address their needs. For future work, this can be achieved with prompt-based AI text generation, e.g. using an OpenAI API.

At the end of this pipeline, the agent outputs a set of enriched leads with scores, reasoning, and a drafted message for each – ready for review by the sales team. The entire process can be run automatically via a single script (see instructions on git), producing both [CSV files](#) and an interactive **dashboard** for visualization.

Some demo in the dashboard (using synthetic data).

Interactive dashboard for DuPont Tedlar's Graphics & Signage sales team to prioritize and manage leads.

Lead Prioritization

Below are the prioritized leads for DuPont Tedlar's Graphics & Signage team, ranked by lead score.

Show 10 entries

Search: avery

Stakeholder	Title	Lead Score	Tier	Company	Industry	Company Size	Email
FirstName3 LastName3	Project Manager	0.84	Tier 1	Avery Dennison Graphics Solutions	Graphics & Signage	Medium	firstname3.lastname3@graphics.averydennison.com
FirstName1 LastName1	Facility Manager	0.86	Tier 1	Avery Dennison Graphics Solutions	Graphics & Signage	Medium	firstname1.lastname1@graphics.averydennison.com
FirstName2 LastName2	Creative Director	0.89	Tier 1	Avery Dennison Graphics Solutions	Graphics & Signage	Medium	firstname2.lastname2@graphics.averydennison.com
FirstName4 LastName4	Sales Director	0.89	Tier 1	Avery Dennison Graphics Solutions	Graphics & Signage	Medium	firstname4.lastname4@graphics.averydennison.com

Showing 1 to 4 of 4 entries (filtered from 51 total entries)

Previous

1

Next

Lead Overview

Lead Distribution by Tier

A pie chart illustrating the distribution of leads across three tiers. The chart is entirely green, representing Tier 1, which accounts for 100% of the leads. A legend to the right shows Tier 1 in green, Tier 3 in red, and Tier 2 in yellow.

Tier	Percentage
Tier 1	100%
Tier 2	0%
Tier 3	0%

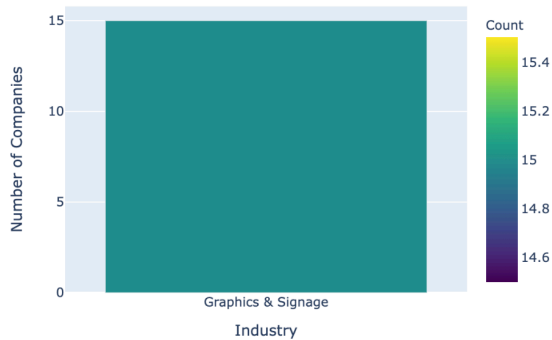
Lead Score Distribution

A histogram showing the distribution of lead scores. The x-axis represents the Lead Score, ranging from 0.85 to 1.0. The y-axis represents the Number of Leads, ranging from 0 to 10. The distribution is as follows:

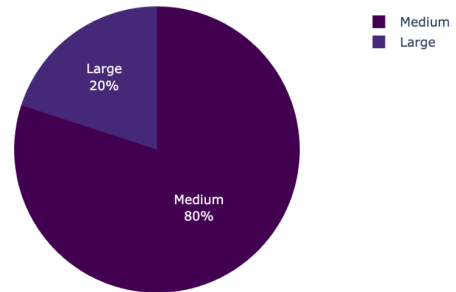
Lead Score Range	Number of Leads
0.85 - 0.875	4
0.875 - 0.9	10
0.9 - 0.925	9
0.925 - 0.95	6
0.95 - 0.975	7
0.975 - 1.0	2
1.0 - 1.025	1

Company Insights

Companies by Industry

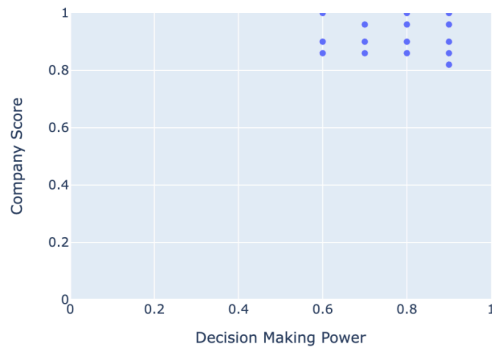


Companies by Size

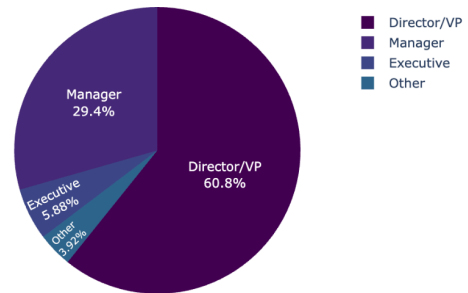


Stakeholder Analysis

Stakeholder Decision Power vs Company Score



Stakeholders by Job Level



Data Processing Steps and Module Overview

The prototype is organized into modular Python components, each handling a phase of the pipeline. The [run_pipeline.py](#) script orchestrates these modules in sequence, passing data from one step to the next. Below we describe each major module and how they interact:

Event Scraping (event_scraper.py & isa_expo_scraper.py)

EventScraper is responsible for finding relevant industry events and extracting participating companies. It contains predefined targets like ISA Sign Expo, PRINTING United, FESPA, etc., and includes custom logic for each event website. In particular, the **ISAEexpoScraper** sub-module was implemented to scrape the ISA Sign Expo 2025 site for its exhibitor list (companies, booth info, descriptions, etc.). In practice, scraping the live ISA site can be complex and time-consuming, so for this prototype the exhibitor data was **seeded with a static list** of

real ISA 2025 companies. This ensured we had real company names and descriptions without relying on real-time web scraping during the demo.

Example: The system pulls **ISA Sign Expo 2025** as a target event. Using the `ISAEexpoScraper` (or the pre-scraped list in [isa_expo_companies.py](#)), it obtains companies like *Avery Dennison Graphics Solutions* (booth 3837) with a description “*provides materials and solutions for graphics applications*”. These companies are labeled with the event source and given an initial relevance score based on keyword matching (e.g. presence of “graphics” or “vinyl” in their profile). The output of this stage is a DataFrame (and CSV) of event-derived companies.

Association Scraping (association_scraper.py)

AssociationScraper collects data on **industry associations** related to printing, signage, and graphics. The module holds a list of relevant associations (e.g. International Sign Association, Specialty Graphic Imaging Association, PRINTING United Alliance, etc.) along with their websites and a subjective relevance score. In the current prototype, this primarily serves to document what groups exist and could be leveraged; the scraper fetches each association’s page to possibly find additional info (such as membership size or upcoming events), but **does not yet extract member company lists**. (Fully parsing association membership directories was beyond the initial scope and would be a future enhancement.) The associations data is saved to a CSV for reference. While association data didn’t directly feed into lead generation due to time constraints, it demonstrates how the system could broaden lead sourcing beyond trade shows in the future.

Company Data Aggregation (company_scraper.py)

The **CompanyScraper** module is responsible for merging leads from events and associations into a unified company list. In this prototype, since the primary source was ISA Sign Expo, the `collect_companies_data` function simply takes the companies identified from the ISA expo (via the static dataset of exhibitors) and returns them as the master list of target companies. This list, saved as [companies.csv](#), includes each company’s name, website, a brief description, the source (event or association), and an initial relevance indicator. If multiple events/associations were used, this module would also handle de-duplicating companies and combining their info. I used a helper to generate **sample companies** for testing, illustrating how one might inject dummy companies if no real data were available.

Notably, real exhibitor data from a major event was used (rather than fully fictional companies) to ground the prototype in reality. However, companies outside the expo or from association directories were not added due to time, so the list is not exhaustive—just sufficient to simulate the pipeline end-to-end.

Company Enrichment & Lead Scoring (company_enricher.py)

Once the raw company list is prepared, the **CompanyEnricher** module enriches each company’s data to facilitate lead qualification. This involves augmenting the company profiles with additional attributes and computing a lead score. Key enrichment steps include:

- **Industry Classification:** Examining the company's description and website for keywords to determine what industry segments apply (e.g. signage, graphics, manufacturing, materials). If the company profile already had an industry listed (in this case, all expo companies were known to be in Graphics/Signage), that is used; otherwise, keywords like "print, wrap, vinyl, etc." are counted. For example, a company with many signage-related terms might be classified under "*signage*" industry.

- **Company Size Estimation:** For future work, if external data were available (e.g. employee count or revenue from an API), size is categorized (Micro, Small, Medium, Large). In absence of real data, a heuristic uses any clues in the description (mentions of "global" or "family-owned", etc.) to guess size. Many expo companies ended up defaulting to "Medium" in the prototype due to limited info.

- **Product/Material Extraction:** The enrich module parses the description text for mentions of products or materials of interest. For instance, if a description mentions "**laminating solutions, vinyl films, banners, displays**", those keywords are collected. This helps identify what the company produces or uses. Materials like *vinyl*, *PVC*, *film* are especially relevant to DuPont Tedlar (since Tedlar film could be a complementary or competing material).

- **Target Markets:** Similarly, it looks for clues about the markets the company serves (e.g. retail signage, outdoor advertising, vehicle wraps, etc.). This gives context on where the company's products are used.

- **Relevance Score Calculation:** Using all the above enriched info, a numeric **lead score** (0 to 1) is calculated for each company reflecting its fit for DuPont Tedlar's Graphics & Signage business. The scoring algorithm in the prototype is a simple rule-based formula: it adds points for desirable traits like operating in the signage/graphics industry, making relevant products (signs, billboards, wraps), using relevant materials (vinyl, films), serving key markets (outdoor advertising, retail, events), and being of a certain size (larger companies score slightly higher as potential high-value customers). Each factor contributes up to 1 point, and the total is normalized to 0–1. For example, a company that prints large-format vinyl banners and serves outdoor advertisers might score **0.9**, whereas a general graphic design studio might score **0.5**. This score is used to **rank and tier the leads** – e.g. Tier 1 (high priority) for scores above 0.8, Tier 2 for 0.5–0.8, etc., in concept.

All enriched company data and scores are output as [enriched_companies.csv](#). Additionally, the enriched data provides a basis for explaining *why* a lead is qualified – e.g. "*Company uses vinyl banner materials (a key use case for Tedlar films) and operates in outdoor signage (Tedlar's target application)*". These rationales can be shown in the dashboard under each company's details, giving the sales team transparency into the lead scoring (future work).

(*Synthetic data usage:* The enricher is designed to integrate real data via APIs like **Clearbit** or Crunchbase to pull firmographics automatically. The code includes a stub for Clearbit enrichment – which would fetch actual industry, employee count, etc. (For future work, if an API

key is provided). Thus, the enrichment relies entirely on local keyword heuristics, and fields like exact employee counts remain blank or estimated.)*

Stakeholder Identification (`stakeholder_finder.py`)

Having prioritized company leads, the next step is to find specific people at those companies to contact. The **StakeholderFinder** module handles this by searching for key job titles commonly associated with decision-makers in relevant domains (executives, marketing heads, operations managers, etc.). The process in the prototype is two-fold:

- **API-based search (future work, optional):** The system is set up to use the **Hunter.io** API (an email and contact finder service) if an API key is provided. For each company, if a website domain is known (e.g. *averydennison.com*), the code would query Hunter's domain search to retrieve publicly available emails and names of people at that company, then filter those by relevant titles (like *Marketing Director* or *VP of Sales*). This could yield real stakeholder names and emails. However, for this demo no live API was used (no key provided), so this branch is essentially skipped.
- **Synthetic stakeholder generation:** In absence of real contact data, the module **simulates stakeholders** to populate the pipeline. For each company, it randomly generates 2–4 fake contacts with plausible titles. It uses a predefined dictionary of relevant job titles (grouped by category: executive, marketing, operations, purchasing, technical, sales) and picks a few categories to assign. Each synthetic stakeholder gets a placeholder name (e.g. "*FirstName1 LastName1*" for the first contact, etc.), an email address constructed from the company's domain (or name if domain unknown), and their role (e.g. "*Marketing Manager*" or "*COO*"). The code also assigns a rough "*decision_making_power*" score based on the seniority of the title (for instance, a CEO might get 0.9 vs. a Sales Manager 0.6) to indicate how influential that contact likely is. While these contacts are not real, they allow the pipeline to continue as if we had identified actual people.

The output is a [scored_leads.csv](#) listing all identified (or generated) stakeholders per company, including fields like Stakeholder ID, name, title, email, company name, `lead_score`, tier, `company_score` and source. This would normally be where a real system integrates with tools like LinkedIn – in a production setting, one would replace the dummy contacts with actual individuals found via Sales Navigator or a similar service. Nonetheless, for each of our target companies we now have at least a placeholder "*person*" to whom an outreach can be addressed.

Personalized Message Generation (`message_generator.py`)

With a target stakeholder and an enriched company profile, the final step is to craft a personalized outreach message. The **MessageGenerator** module uses the company and stakeholder data to produce a tailored message, aiming to spark the stakeholder's interest in DuPont Tedlar's products. For future work, this would likely leverage an **OpenAI GPT model** (or other LLM) by feeding a prompt that includes details like the stakeholder's name, title, company

background, and key value propositions of DuPont Tedlar for that context. The model would then generate a custom email or LinkedIn message text.

In the current prototype, the message generation is simplistic (a placeholder, since full prompt engineering was marked as future work). It likely follows a template such as:

"Hello {Stakeholder Name}, I noticed your role as {Title} at {Company}. {Company}'s focus on {highlight from company profile} caught my eye – it aligns closely with solutions we provide at DuPont Tedlar in the graphics and signage space. I'd love to discuss how our Tedlar films could help {Company} achieve {some relevant benefit}..."

Each stakeholder is thus associated with a brief outreach message string, which is appended to their record. The results are compiled into [stakeholders_with_messages.csv](#) – essentially an enriched stakeholder list now including the AI-generated message for each. These messages are meant to be starting points, requiring review and refinement by the sales team, but they demonstrate how the AI agent can automate the first draft of personalized outreach at scale.

(Note: In this prototype, actual calls to the OpenAI API were not made due to time and cost constraints. The focus was on setting up the pipeline such that plugging in a model is straightforward in the future. For future work, we should consider prompt optimization, tone adjustments, and ensuring factual correctness in the message are acknowledged as needed iterations once real generation is enabled.)

Orchestration & Dashboard (run_pipeline.py and Output)

All the above components are integrated in run_pipeline.py, which runs each step sequentially and passes its output to the next. After generating outreach messages, the script also invokes a **DashboardGenerator** to present the results. The final output of the pipeline is both a set of CSV files as described, and an **interactive dashboard** (simulated in a Streamlit-like format). For future work, the dashboard should present a **lead table** where companies are listed along with their priority tier or score, and the user can expand each company's entry to see details like the company description, the rationale for its qualification (relevance factors), the list of identified stakeholders, and the recommended outreach message for each stakeholder. There may also be simple charts summarizing the lead pool (e.g. count of leads by industry or lead tier).

Now this dashboard allows for quick review and validation of the AI agent's outputs in a user-friendly way. It could be improved by adding more features. For example, a sales engineer can sort the table by relevance score to see top opportunities (Tier 1 leads at the top) or filter by a keyword. Under each company, I will consider adding the **"why qualified"** explanation and the message text, so the user can immediately gauge if the lead makes sense and if the outreach approach is appropriate. Now, the prototype's dashboard is generated as a local HTML file, which can be opened in a browser for interaction.

Dashboard Screenshot of Lead Table:

Lead Prioritization

Below are the prioritized leads for DuPont Tedlar's Graphics & Signage team, ranked by lead score.

Show 10 entries
Search:

Stakeholder	Title	Lead Score	Tier	Company	Industry	Company Size	Email
FirstName2 LastName2	Marketing Director	1.0	Tier 1	Lintec of America, Inc.	Graphics & Signage	Large	firstname2.lastname2@lintecofamerica.com
FirstName3 LastName3	Director of Purchasing	1.0	Tier 1	Lintec of America, Inc.	Graphics & Signage	Large	firstname3.lastname3@lintecofamerica.com
FirstName1 LastName1	Technical Director	1.0	Tier 1	General Formulations	Graphics & Signage	Large	firstname1.lastname1@generalformulations.com
FirstName3 LastName3	CFO	1.0	Tier 1	General Formulations	Graphics & Signage	Large	firstname3.lastname3@generalformulations.com
FirstName4 LastName4	Director of Operations	1.0	Tier 1	Lintec of America, Inc.	Graphics & Signage	Large	firstname4.lastname4@lintecofamerica.com
FirstName1 LastName1	Sales Director	1.0	Tier 1	Lintec of America, Inc.	Graphics & Signage	Large	firstname1.lastname1@lintecofamerica.com

Implementation Results (included ideal future work output)

By running the pipeline, we obtain a **prioritized list of leads** enriched with data and AI-generated insights. In summary, the system outputs:

- Enriched Company Leads:** A list of ~50+ companies (exhibitors from ISA Sign Expo 2025 in this prototype) each annotated with industry tags, estimated size, relevant products/materials, target markets, and a calculated relevance score (0.00–1.00). Based on the score, each company can be categorized into tiers (e.g. Tier 1 for top prospects). The enrichment process also provides context for qualification — for instance, a company might have *“Relevance 0.88 (High) – manufactures vinyl banners and outdoor signage (key Tedlar use cases)”* as a note.
- Stakeholder Contacts:** For each company, 2–4 stakeholder contacts are listed, including names (actual or placeholder), job titles, and emails. While the demo uses synthetic names, it shows the roles that would be targeted (e.g. *CEO, VP of Marketing, Operations Manager* at that company). This verifies that the agent can supply at least one appropriate contact per lead to reach out to.
- Personalized Outreach Messages:** A short, custom-crafted message for each stakeholder, referencing their company’s specifics. The messages are output as part of the stakeholder list. For example, one message might read: *“Hi Jane Doe, as the Marketing Director at XYZ Signs, I thought you’d be interested in how DuPont’s Tedlar films can enhance the durability of your outdoor graphics…”*. Each message is unique to the company’s scenario, illustrating a level of personalization that would be time-consuming to do manually for dozens of leads.
- Exportable CSV files:** All the above information is saved into CSVs (events.csv, associations.csv, companies.csv, enriched_companies.csv, stakeholders.csv, stakeholders_with_messages.csv) for record-keeping or further analysis. A sales team could

import these into a CRM or spreadsheet. The CSVs also make it transparent what data was gathered or assumed at each step.

- **Interactive Dashboard (Demo UI):** As described, the prototype includes a simple web-based dashboard to explore the leads. In a live setting, this could be shown to stakeholders to demonstrate the system's outputs. The dashboard's interactive elements (sorting, expanding rows, etc.) help in quickly scanning which leads are high priority and examining the details and messages without sifting through raw CSVs.

Overall, the result of the pipeline is a **proof-of-concept deliverable**: it shows that starting from open-source data (events and associations), one can automate the generation of a targeted outreach list complete with messaging. The content is prepped for a go-to-market (GTM) team to review. Importantly, because this is a prototype, the outputs are not yet truly actionable (e.g. contact info is fake, messaging is draft quality), but they paint a clear picture of how a fully implemented system could save enormous time in lead research and initial outreach formulation.

Future Work and Improvements

There are many practical enhancements that could turn this prototype into a production-ready tool. Given more time and resources, I will consider the following improvements:

- **Real Stakeholder Lookup – LinkedIn & Clay Integration:** Incorporate services like **LinkedIn Sales Navigator API** or **Clay** to fetch actual people at target companies. This would replace synthetic contacts with real names, titles, and even direct LinkedIn URLs or phone numbers. For example, the system could automatically find the *Marketing VP at Avery Dennison* and pull their profile details, greatly increasing the usefulness of the leads list.

- **Advanced Company Enrichment via APIs:** Leverage external data platforms such as **Crunchbase** or **Clearbit** to enrich company information. These APIs can provide up-to-date firmographic data (industry classifications, employee counts, revenue ranges, key people, etc.). In a future version, the script would call these services for each company to get more accurate data instead of relying on keyword guesses. This would improve lead scoring accuracy and fill in missing details (like company size or location).

- **Prompt Optimization for AI Messaging:** Once real data is in place, invest in **prompt engineering** for the OpenAI (or other LLM) that generates the outreach messages. This could involve providing the model with a structured prompt containing the company's profile, the stakeholder's role, and specific value propositions of Tedlar for that scenario. Testing and refining the prompt will yield more compelling and correct messages. Additionally, one could fine-tune a model on successful sales emails in the graphics/signage industry to further improve relevance and tone. The goal is to ensure the AI's messages are high-quality and require minimal editing.

- **Automated Pipeline Execution:** Set up the pipeline to run on a schedule or trigger so that lead generation becomes a continuous process. For instance, using a **cron job**

or GitHub Actions, the system could scrape new events or updates weekly and refresh the lead list. This automation would keep the sales pipeline topped up with minimal manual effort. Integration with CI/CD also ensures the latest code and data are used, and any issues can be caught via automated tests.

- **Persistent Database and History:** Instead of (or in addition to) CSV outputs, use a **database** to store leads and enrichment results. A relational database (like PostgreSQL) or a cloud CRM database would allow persistent storage of leads over time, tracking which leads have been contacted or which events they came from. This would make the tool more robust for long-term use, enabling queries like “show new leads added this month” or avoiding duplicate outreach to the same company from a past event.

- **Dashboard UI Enhancements:** Improve the front-end for reviewing leads. A next version of the dashboard could be a true **web app** (possibly built with Streamlit or a React frontend) that allows filtering by score range, searching by company name, and editing or approving the AI-generated messages directly in the interface. Features like exporting selected leads to CSV/Excel, or marking a lead as “qualified”/“not interested”, would make it a functional tool for a GTM team. The design can also be polished with charts (e.g. a bar chart of lead count by industry or a pie chart of lead tiers) to give a quick analytical overview.

- **CRM and Email Integration:** To close the loop, the system can be connected with **CRM software** (like Salesforce or HubSpot) and email outreach tools. For example, an approved lead and its contact could be pushed into Salesforce with one click, or the personalized email text could be fed into an email automation platform to send directly to the stakeholder. This would turn the prototype into a fully automated lead gen + outreach engine, where the only human task is to supervise and refine the AI’s output.

- **Robust Scraping & Additional Sources:** Expand the scraping capability to cover more events (international signage expos, printing conferences) and perhaps supplier directories or online marketplaces in the signage industry. Each additional source can feed new leads into the pipeline. This will require making the web scrapers more robust (handling different site structures, login requirements for member lists, etc.) and possibly using search engine APIs to discover new leads continually. With more time, the event and association scrapers could be enhanced to automatically traverse known sites and even find **new** industry events or organizations.

In conclusion, this case study prototype demonstrates the potential of an AI-assisted GTM workflow: it can take unstructured web data and transform it into a curated set of sales leads with personalized outreach content. While currently a simplified **proof-of-concept**, with further development – integrating real data sources, improving AI outputs, and building automation around it – this system could significantly accelerate the lead generation and initial contact process for DuPont Tedlar’s team (or any similar B2B sales organization).