Cloud deployment

The architecture can be viewed in the diagram, so instead of describing it again I will explain what's the purpose of each service we used:

- Auto Scaling Group (ASG) for seamless scalability, our backend will scale based on demand / server health. (and CloudWatch metrics which we will dive in later)
- Application Load Balancer (ALB) for distributing traffic evenly across our instances so that they are not overloaded + a stable entry point for the application (it never changes, unlike the EC2 instances behind it)
- Relational Database Service (RDS) for our PostgreSQL database. We chose it because it is a managed service and gives us less headaches. It has automatic patches from AWS and minor version upgrades, as well as replication and encryption.
- CloudFront for caching at the edge, leveraging AWS's Edge Locations so that we cache content as close to the user as possible, ensuring lower latency
- Certificate Manager (ACM) for encryption in flight, storing out SSL certificates used for HTTPs.
- Route53 is the registrar of our domain + DNS server that manages it (Hosted Zone).
- S3 is where our website is stored (the static files of it). Cheap and durable storage, ideal for our situation. It only allows CloudFront as an entrypoint so basically HTTPs is enforced, the only access to the bucket is through the CloudFront OAC.
- SSM (Systems Manager) is used for storing parameters securely in the Parameter Store, from where EC2 instances assume an IAM role which allows them to take the credentials to the database.
- SNS is integrated with our ASG in the backend, such that admins receive notification on EC2 Instance launches / terminations or other notifications.
- API Gateway fronts our Lambda Function, integrating together seamlessly for a fully serverless service combination, that's where our AI Receipt Analyzer sits at. The API Gateway is there for ease of access via HTTP since Lambda URLs are not recommended. The intention behind the API Gateway is also so that it can be called from the outside (i.e. the mobile application, since if it was only called from the EC2 backend instances then we would have skipped the API Gateway and just invoked the Lambda right away)
- KMS is storing our encryption keys with which we encrypt information in our cloud such as the parameters in SSM Parameter Store, etc. It has been explicitly chosen for its automatic rotation of keys.
- CloudWatch is taking logs from our Lambda function and we are using a CloudWatch Metric to monitor the CPU usage of our backend, keeping it on average below 50%.
- GuardDuty (currently disabled) is therefore using ML to analyze traffic in our VPC (DNS traffic, CloudTrail logs etc.) for suspicious activity.
- Web Application Firewall (WAF) for Layer 7 exploits protection (XSS, SQL Injection etc.)
- Shield is there for Layers 3 and 4 protection (DDoS protection)